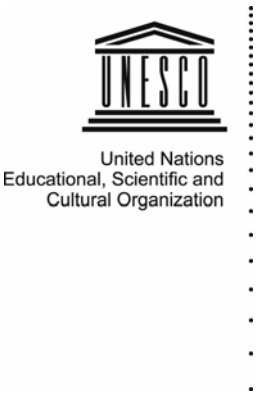


2005/ED/EFA/MRT/PI/44

Background paper prepared for the

Education for All Global Monitoring Report 2005

The Quality Imperative

Review of school and instructional effectiveness research

Jaap Scheerens

2004

REVIEW OF SCHOOL AND INSTRUCTIONAL EFFECTIVENESS RESEARCH

CONTRIBUTION TO CHAPTER 3 OF THE 2004 EFA GLOBAL MONITORING REPORT

Jaap Scheerens, March, 2004

In this chapter an overview will be given of the research literature on educational effectiveness research. The term "educational effectiveness" is used as a general term that encompasses school and instructional effectiveness. "School effectiveness" refers to effectiveness enhancing conditions defined at school level and "instructional effectiveness" to effectiveness enhancing conditions situated at the teacher and classroom level. Multi-level definitions, in which school level conditions, classroom level conditions and usually also conditions in the larger context of the school are included are sometimes referred to as "integrated school effectiveness models" and sometimes as "integrated educational effectiveness models".

In the first part of the chapter school effectiveness and integrated educational effectiveness studies are reviewed, while the second part of the chapter is totally dedicated to instructional effectiveness. Instead of "instructional effectiveness" terms like teacher and teaching effectiveness are also used.

PART I SCHOOL EFFECTIVENESS AND INTEGRATED EDUCATIONAL EFFECTIVENESS MODELS

The overall design of educational effectiveness studies

The elementary design of school effectiveness research is the association of hypothetical effectiveness enhancing conditions of schooling and output measures, mostly student achievement. The basic model from systems theory that was introduced in chapter one is helpful to clarify this basic design.(see Figure 1). The major task of school effectiveness research is to reveal the impact of relevant input characteristics on output and to "break open" the black box in order to show which process or throughput factors "work", next to the impact of contextual conditions. Within the school it is helpful to distinguish a school and a classroom level and, accordingly, school organizational and instructional processes.

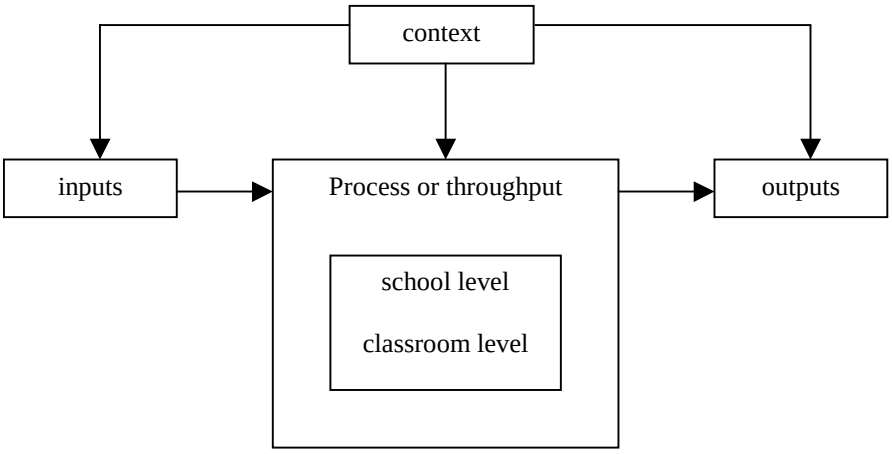


Figure 1: A basic systems model on the functioning of education

Research tradition in educational effectiveness varies according to the emphasis that is put on the various antecedent conditions of educational outputs. These traditions also have a disciplinary basis. The common denominator of the five areas of effectiveness research that will be distinguished is that in each case the elementary design of associating outputs or outcomes of schooling with antecedent conditions (inputs, processes or contextual) applies. The following research areas or research traditions will be considered in summarizing the research results obtained in developed countries:

1) ***Research on equality of opportunities in education and the significance of the school in this.***
2) ***Economic studies on education production functions.***
3) ***The evaluation of compensatory programs.***
4) ***Studies of unusually effective schools.***
5) ***Studies on the effectiveness of teachers, classes and instructional procedures.***

In developing countries there is a strong predominance of studies of the education production function type. Relatively few of these have been expanded by including school organizational and instructional variables.

Results obtained in various strand of educational effectiveness research

re 1) School effectiveness in equal educational opportunity research

Coleman's research into educational opportunity, about which a final report known as the Coleman report was published in 1966, forms the corner-stone for school effectiveness studies (Coleman et al., 1966). While this study was intended to show the extent to which school achievement is related to students' ethnic and social background, the possible influence of the "school" factor on learning attainment was also examined.

In the survey three clusters of school characteristics were measured: (a) teacher characteristics; (b) material facilities and curriculum; and (c) characteristics of the groups or classes in which the pupils were placed. After the influence of ethnic origin and socio-economic status of the pupils had been statistically eliminated, it appeared that these three clusters of school characteristics together accounted for 10 percent of the variance in pupil performance. Moreover, the greater part of this 10 percent variance was due to the third cluster that was operationalized as the average background characteristics of pupils, which means that again the socio-economic and ethnic origin - now defined at the level of the school - played a central role. In reactions to the Coleman report there was general criticism on the limited interpretation of the school characteristics. Usually, only the material characteristics were referred to, such as the number of books in the school library, the age of the building, the training of the teachers, their salaries and expenditure per pupil. Nevertheless there were other characteristics included in Coleman's survey, such as the attitude of school heads and teachers towards pupils and the attitude of teachers towards integrated education, i.e. multiracial and classless teaching.

Other large-scale studies that were primarily focused at providing data on equality of opportunity are those by Jencks et al. (1972, 1979), Alexander and Eckland (1980), and Hauser, Sewell and Alwin (1976). Thorndike's (1973) study, although not explicitly dedicated to equality of opportunity, also examined school careers in relationships to the environmental background of pupils.

The overall results of these studies indicated a relatively high correlation between socio-economic and ethnic family characteristics and learning attainment, and a small or even negligible influence from school and instruction characteristics. The outcomes were criticized by educationalists for the rather narrow choice of school characteristics and on methodological grounds (cf. Aitkin & Longford, 1986), for multi-level associations not being properly modelled and analyzed.

re 2) Economic studies on educational production functions

The focus of economic approaches towards school effectiveness is the question of what manipulative inputs can increase outputs. If there was stable knowledge available on the extent to which variety of inputs is related to variety of outputs it would also be possible to specify a function which is characteristic of the production process in schools. Stated differently: a function, which could accurately indicate how a change in the inputs would affect the outputs.

This leads to a research-tradition that is identified both by the term input-output studies as by the term research into education production functions. The research model for economics-related production studies hardly differs from that for other types of effectiveness research: the relationship between manipulative school characteristics and attainment is studied while

the influence of background conditions like social class and pupils' intelligence is eliminated as far as possible. The specific nature of production-function research is the concentration on what can be interpreted in a more literal sense as input characteristics: the teacher/pupil relationship, teacher training, teacher experience, teachers' salaries and expenditure per pupil. In more recent observations of this research type one comes across the suggestion to take effectiveness predictors known from educational psychology research into account (Hanushek, 1986). It should be noted that the Coleman-report (Coleman et al.1966) is often included in the category of input-output studies. In view of its emphasis on the more material school characteristics, the association is an obvious one.

The findings of this type of research have often been referred to as being disappointing. Review studies like those from Mosteller and Moynihan (1972), Averch et al. (1974), Glasman and Biniaminov (1981), Hanushek (1979 and 1986) always produce the same conclusions: inconsistent findings throughout the entire available research and scant effect at most from the relevant input variables.

From reanalysis of Hanushek's (1986) dataset, Hedges et al. (1994), however, conclude that there is an effect of per pupil expenditure of "considerable practical importance" (an increase of PPE by $510 would be associated with a 0.7 s.d. increase in student outcome). But this conclusion in its turn is contested by Hanushek. In Table 1 cited from Hanushek, 1997, the most recent "vote count" overview of education production function studies is given.

Table 1: *Percentage Distribution of Estimated Effect of Key Resources on Student performance, Based on 377 Studies (cited from Hanushek, 1997, p. 144)*

		Statistically significant		Statistically insignificant		
Resources	Number of estimates	Positive	Negative	Positive	Negative	Unknown sign
Real classroom resources						
Teacher pupil ratio	277	15%	13%	27%	25%	20%
Teacher education	171	9	5	33	27	26
Teacher experience	207	29	5	30	24	12
Financial aggregates						
Teacher salary	119	20%	7%	25%	20%	28%
Expenditure per pupil	163	27	7	34	19	13

Hanushek's interpretation of these results is that there can be little confidence that adding more of any of the specific resources or, for that matter of the financial aggregates, will lead to a boost in student achievement. The variable that shows relatively the highest proportion of positive effects is teacher experience, but here, "reverse causation" could be at play, since more experienced teachers might have selected schools with better performing pupils (ibid, p. 144).

In other reviews, e.g. Verstegen & King (1998), a more positive interpretation is given on largely the same set of studies that was analyzed by Hanushek (1997). During the last decade several studies drew attention to the fact that certain resource input factors did show significant positive associations with pupil achievement or other educational outcomes. The most important of these are the studies by Card & Krueger (1992), which indicated a positive association between school resources and differences in earnings among workers, Hedges, Laine & Greenwald (1994) who conducted a statistical meta-analysis on a sub-set of Hanushek's 1979 data set and found significant effects for several resource input variables, among which is rather large positive effect of Per Pupil Expenditure, Ferguson (1991), who found particularly large effects of variables related to teacher qualifications (specifically scores on a teacher recertification test), and Achilles (1996) who reported the sustained effects of reduced class-size (14-16 as compared to 22-24) in Kindergarten and the first three grades of primary school) on student achievement.

That these differences in interpretation are to a certain degree of the kind: "the cup is half full" as compared to "the cup is half empty" is illustrated by Verstegen & King's (1998) presentation of table 6, cited from Hanushek, 1997.

Table 2: *Verstegen & King's (1998) rendering of Hanushek's (1997, p. 144) tabulation.*

Percentage Distribution of Significant Estimated Effects of Key Resources on Student Achievement, Based on 377 Studies			
		Statistically significant	
	Number of Estimates (no.)	Positive (%)	Negative (%)
Real Classroom Resources			
Teacher-pupil ratio	78	54	46
Teacher education	24	64	36
Teacher experience	70	85	15
Financial aggregates			
Teacher salary	32	74	26
Expenditure per pupil	55	79	21

By omitting the large proportions of studies showing insignificant results, and "blowing up" the relatively small numbers of studies showing significant results to percentages, these authors appear to be keen to see (or construct) the bright side of things.

Unfortunately, as in other types of educational effectiveness studies, the critics and those who present the more conservative interpretation appear to have the best arguments. Hanushek, 1997, presents most of them: when outcome measures, such as student achievement scores are properly adjusted for student background characteristics, and "value added" outcome indicators are used, the number of positive effects declines. If data at high aggregation levels (e.g. individual states) is used misspecification bias is likely to produce overstatement of effects (this criticism would apply to both the Ferguson and Card & Krueger studies). This problem frequently occurs for the variable Per Pupil Expenditure which is usually only defined at the district level. In statistical meta-analysis the null-hypothesis that is addressed is that resources or expenditure differences never, under whatever circumstances, affect student performance; clearly this hypothesis is to be rejected also in cases where only a minority of studies shows a significant positive association with the outcome variable.

Many of the recent contributions to summarizing the research evidence on education production function studies mention the need to search for answers to the question "why money does or does not matter", for example by looking for combinations and interactions between resource input levels and school organizational and instructional variables. In a recent collection of articles on class size (Galton, 1999) reference is made to differences between educational cultures in the degree to which large classes are considered a burden to teachers.

Another desirable extension of the basic education production function type of study would be to address questions of cost-effectiveness more directly, by comparing cost-effectiveness or even cost-benefit ratio's for different policy measures. A comparison of education production function studies between industrialized and developing countries is particularly interesting, since a "restriction of range" phenomenon (little variance in, for example, teacher salaries between schools) might surpress the effects in relatively homogenous school systems. Results of education production function studies in developing countries will be presented in a subsequent section.

re 3) The evaluation of compensatory programs

Compensatory programs may be seen as the active branch in the field of equal educational opportunity. In the United States compensatory programs like "Head Start" were part of President Johnson's "war on poverty". Other large-scale American programs were "Follow-Through" - the sequel to Head Start - and special national development programs that resulted from Title 1 of the Elementary and Secondary Education Act, enacted in 1965. Compensatory programs were intended to improve the levels of performance of the

educationally disadvantaged. In the late sixties and early seventies there were also similar programs in the Netherlands like the Amsterdam Innovation project, the Playgroup Experiment project, Rotterdam's Education and Social Environment (OSM) project and the Differentiated Education project (GEON) of the city of Utrecht. Compensatory programs manipulate school conditions in order to raise achievement levels of disadvantaged groups of pupils. The level in which this is achieved demonstrates the importance of the school factor - and in particular the conditions and educational provisions within it. However, it proved to be not that simple to redress the balance with effective compensatory programs. In fact no overwhelming successes could be established. There was heated debate on the way available evaluation studies should be interpreted.

The key question is: what results can be realistically expected from compensatory education given the dominant influence in the long run of family background and cognitive aptitudes on pupils' attainment level? Scheerens (1987, p. 95) concluded that the general image provided by the evaluation of compensatory programs reveals that relatively small progress in performance and cognitive development can be established immediately after a program finishes. Long-term effects of compensatory programs cannot be established by and large. Moreover, it has been occasionally demonstrated that it was the "moderately" disadvantaged in particular that benefited from the programs, while the most educationally disadvantaged pupils made the least progress, relatively speaking.In view of the variety of compensatory programs the evaluation studies gave some insight into the relatively best type of educational provision. When comparing the various components of Follow Through, programs aimed at developing elementary skills like language and mathematics and which used highly structured methods turned out to be winners (Stebbins et al., 1977; Bereiter & Kurland, 1982; Haywood, 1982).

As will appear later, there is a remarkable similarity between these characteristics and the findings of other types of effectiveness research. In any case, when interpreting the results of evaluations of compensatory programs one should be aware that the findings have been established among a specific pupil population: very young children (infants or first years of junior school) from predominantly working-class families.

re 4) Effective schools research

Research known under labels like "identifying unusually effective schools" or the "effective schools movement" can be regarded as the type of research that most touches the core of school effectiveness research. In Coleman's and Jencks' surveys the inequality of educational opportunity was the central problem. In economic-related input-output studies the school was even conceived as a "black box". In the still to be discussed research on the effectiveness of classes, teachers and instruction methods, education characteristics on a lower aggregation level than the school are the primary research object.

Effective school research is generally regarded as a response to the results of studies like Coleman's and Jencks' from which it was concluded that schools did not matter very much when it came down to differences in levels of achievement. From titles such as "Schools can make a difference" (Brookover et al., 1979) and "School matters" (Mortimore et al., 1988) it appears that refuting this message was an important source of inspiration for this type of research. The most distinguishing feature of effective schools research was the fact that it attempted to break open the "black box" of the school by studying characteristics related to organization, form and content of schools. The results of the early effective schools research converged more or less around five factors:

strong educational leadership;

emphasis on the acquiring of basic skills;

an orderly and secure environment;

high expectations of pupil attainment;

frequent assessment of pupil progress.

In the literature this summarizing is sometimes identified as the "five-factor model of school effectiveness". It should be mentioned that effective schools research has been largely carried out in primary schools, while at the same time studies have been largely conducted in inner cities and in predominantly working-class neighborhoods.

In more recent contributions effective schools research became more integrated with education production function and instructional effectiveness research, in the sense that a mixture of antecedent conditions was included, studies evolved from comparative case-studies to surveys and conceptual and analytical multi-level modeling took place to analyze and interpret the results. Numerous reviews on school effectiveness have been published since the late seventies. Early reviews are those by Anderson (1982), Cohen (1982), Dougherty (1981), Edmonds (1979), Murnane (1981), Neufeld et al. (1983), Purkey and Smith (1983), Rutter (1983), Good and Brophy (1986), Ralph and Fenessey (1983), Kyle (1985), and Sweeney (1982). More recent reviews are those by Levine and Lezotte (1990), Scheerens (1992), Creemers (1994), Reynolds et al. (1993), Sammons et al. (1995), and Cotton (1995).

The focal point of interest in the reviews is the "what works" question; typically the review presents lists of effectiveness enhancing conditions. There is a fairly large consensus on the main categories of variables that are distinguished as effectiveness enhancing conditions in the reviews, also when earlier and more recent reviews are compared. Table 3 summarizes the characteristics listed in the reviews by Purkey and Smith (1983), Scheerens (1992), Levine and Lezotte (1990), Sammons et al. (1995), Cotton (1995).

Table 2.3: *Effectiveness enhancing conditions of schooling in five review studies (italics in the column of the Cotton study refers to sub-categories).*

Purkey & Smith, 1983	*Levine & Lezotte, 1990*	*Scheerens, 1992*	*Cotton, 1995*	*Sammons, Hillman & Mortimore, 1995*
Achievement-oriented policy; cooperative atmosphere, orderly climate	Productive climate and culture	Pressure to achieve, consensus, cooperative planning, orderly atmosphere	Planning and learning goals, curriculum planning and development	Shared vision and goals, a learning environment, positive reinforcement
Clear goals on basic skills	Focus on central learning skills		Planning and learning goals *school wide emphasis on learning*	Concentration on teaching and learning
Frequent evaluation	Appropriate monitoring	Evaluative potential of the school, monitoring of pupils' progress	Assessment (district, school, classroom level)	Monitoring progress
In-service training/ staff development	Practice-oriented staff development		*Professional development* collegial learning	A learning organization
Strong leadership	Outstanding leadership	Educational leadership	School management and organization, leadership and school improvement, leadership and planning	Professional leadership
	Salient parent involvement	Parent support	Parent community involvement	Home school partnership
Time on task, reinforcement, streaming	Effective instructional arrangements	Structured, teaching, effective learning time, opportunity to learn	Classroom management and organization, instruction	Purposeful teaching
High expectations	High expectations		Teacher student interactions	High expectations
				Pupil rights and responsibilities
			Distinct-school	

		interactions
		Equity
		Special programs
	External stimuli to make schools effective	
	Physical and material school characteristics	
	Teacher experience	
	School context characteristics	

Consensus is largest with respect to the factors: achievement orientation (which is closely related to "high expectations"); co-operation; educational leadership; frequent monitoring; time, opportunity to learn and "structure" as the main instructional conditions.

Behind this consensus on general characteristics hides considerable divergence in the actual operationalization of each of the conditions. Evidently concepts like "productive, achievement-oriented climate and educational leadership are complex concepts and individual studies may vary in the focus that different elements receive. Scheerens and Bosker (1997, ch. 4) provide an analysis of the meaning of the factors that are considered to work in schooling apparent from the actual questionnaires and scales as used in ten empirical school effectiveness studies. Their summary table, in which the main components of thirteen general factors are mentioned, is cited below as Table 4.

Table 4: *Components of fourteen effectiveness-enhancing factors*

Factors	**Components**
Achievement, orientation, high expectations	clear focus on the mastering of basic subjects high expectations (school level) high expectations (teacher level) records on pupils' achievement
Educational leadership	general leadership skills school leader as information provider orchestrator or participative decision making school leader as coordinator meta-controller of classroom processes time educational/administrative leadership

	counselor and quality controller of classroom teachers initiator and facilitator of staff professionalization
Consensus and cohesion among staff	types and frequency of meetings and consultations contents of cooperation satisfaction about cooperation importance attributed to cooperation indicators of successful cooperation
Curriculum quality/ opportunity to learn	the way curricular priorities are set choice of methods and text books application of methods and text books opportunity to learn satisfaction with the curriculum
School climate	*orderly atmosphere* the importance given to an orderly climate rules and regulations punishment and rewarding absenteeism and drop out good conduct and behaviour of pupils satisfaction with orderly school climate
	climate in terms of effectiveness orientation and good internal relationships priorities in an effectiveness-enhancing school climate perceptions on effectiveness-enhancing conditions relationships between pupils relationships between teacher and pupils relationships between staff relationships: the role of the head teacher engagement of pupils appraisal of roles and tasks job appraisal in terms of facilities, conditions of labour, task load and general satisfaction facilities and building
Evaluative potential	evaluation emphasis monitoring pupils' progress use of pupil monitoring systems school process evaluation use of evaluation results keeping records on pupils' performance satisfaction with evaluation activities

Parental involvement	emphasis on parental involvement in school policy contacts with parents satisfaction with parental involvement
Classroom climate	relationships within the classroom order work attitude satisfaction
Effective learning time	importance of effective learning time monitoring of absenteeism time at school time at classroom level classroom management homework

re 5) Studies on instructional effectiveness

For the current review studies on characteristics of effective teachers, and studies that go under the label of "process-product studies", are the most relevant strands of research on teaching and classroom processes. This latter category of studies was strongly inspired by Carroll's (1963) model of teaching and learning and off-springs of this model, such as the models of mastery learning (Bloom, 1976) and "direct teaching" (e.g. Doyle, 1985). The research results have been reviewed by, among others, Stallings (1985), Brophy and Good (1986), and Creemers (1994) and quantitatively synthesized in meta-analyses by Walberg (1984), Fraser et al. (1987) and Wang, Haertel and Walberg (1993). These latter authors incidentally have also included variables outside the classroom situation, like the student's relationships with peers, and the home environment (e.g. television viewing) in their analyses which they label under the heading of "educational productivity". The main research results are summarized in part II of this chapter.

Integration

Of the five effectiveness-oriented educational research types, which were reviewed, two focused on "material" school characteristics (such as teacher salaries, building facilities and teacher/pupil ratio). The results were rather disappointing in that no substantial positive correlations of these material investments and educational achievement could be established in a consistent way across individual studies. On the basis of more recent studies these rather pessimistic conclusions have been challenged, although methodological critique indicates that the earlier pessimistic conclusions are more realistic. In-depth process studies connected with

large-scale evaluations of compensatory programs pointed out that programs which used direct, i.e. structured, teaching approaches were superior to more "open" approaches. The research movement known as research on exemplary effective schools (or briefly: effective schools research) focused more on the internal functioning of schools than the earlier tradition of input-output studies. These studies produced evidence that factors like strong educational leadership, emphasis on basic skills, an orderly and secure climate, high expectations of pupil achievement and frequent assessment of pupil progress were indicative of unusually effective schools.

Research results in the field of instructional effectiveness are centered around three major factors: effective learning time, structured teaching and opportunity to learn in the sense of a close alignment between items taught and items tested.

Although all kinds of nuances and specificities should be taken into account when interpreting these general results they appear to be fairly robust - as far as educational setting and type of students is concerned. The overall message is that an emphasis on basic subjects, an achievement-oriented orientation, an orderly school environment and structured teaching, which includes frequent assessment of progress, is effective in the attainment of learning results in the basic school subjects.

Table 5 summarizes the main characteristics of the five research traditions.

Table 5: *General characteristics of types of educational effectiveness research*

	independent variable type	*Dependent variable type*	*Discipline*	*main study type*
(un)equal opportunities	socio-economic status and IQ of pupil, material school characteristics	Attainment	Sociology	Survey
Production functions	material school characteristics	achievement level	Economics	Survey
evaluation compensatory programs	specific curricula	achievement level	interdisciplinary pedagogy	quasi-experiment
Effective schools	"process" characteristics of schools	achievement level	interdisciplinary pedagogy	case-study
Effective instruction	characteristics of teachers, instruction, class organization	achievement level	educational psychology	Experiment observation

In recent school effectiveness studies these various approaches to educational effectiveness have become integrated. Integration was manifested in the conceptual modeling and the choice of variables. At the technical level multi-level analysis has contributed significantly to this development. In contributions to the conceptual modeling of school effectiveness, schools became depicted as a set of “nested layers” (Purkey and Smith, 1983), where the central assumption was that higher organizational levels facilitated effectiveness enhancing conditions at lower levels (Scheerens & Creemers, 1989). In this way a synthesis between production functions, instructional effectiveness and school effectiveness became possible. This was done by including the key variables from each tradition, each at the appropriate “layer” or level of school functioning [the school environment, the level of school organization and management, the classroom level and the level of the individual student]. Conceptual models that were developed according to this integrative perspective are those by Scheerens (1990), Creemers (1994), and Stringfield and Slavin (1992). Since the Scheerens model (also cited in chapter 1) was used as the starting point of the meta-analyses described in subsequent sections it is shown in Figure 2.

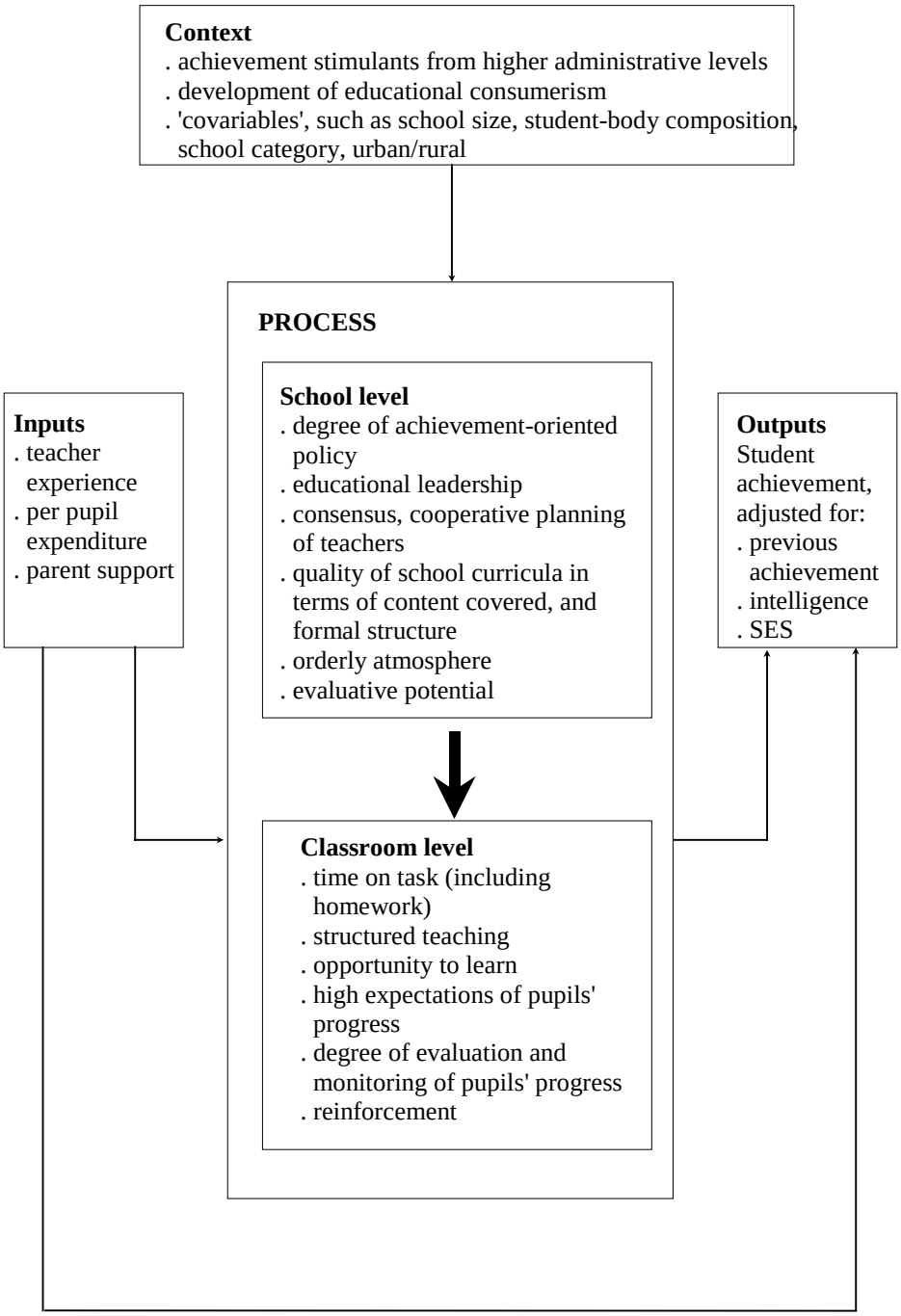


Figure 2.: *An integrated model of school effectiveness (from Scheerens, 1990)*

The choice of variables in this model is supported by the “review of reviews” on school effectiveness research that will be presented in the next section. Exemplary cases of integrative, multi-level school effectiveness studies are those by Mortimore et al. (1988), Brandsma (1993), Hill et al. (1995), Sammons et al. (1995) and Grisay (1996).

Summary of meta-analyses

In Table 6 (cited from Scheerens and Bosker, 1997) the results of three meta-analysis and a re-analysis of an international data set have been summarized. The results concerning resource input variables are based on the re-analysis of Hanushek's (1989) summary of results of production function studies that was carried out by Hedges, Laine & Greenwald, 1994. As stated before this re-analysis was criticized, particularly the unexpectedly large effect of per pupil expenditure. The results on "aspects of structured teaching" are taken form meta-analyses conducted by Fraser, Walberg, Welch and Hattie, 1987. The international analysis was based on the IEA Reading Literacy Study and carried out by R.J. Bosker (Scheerens & Bosker, 1997, ch. 7). The meta-analysis on school organizational factors, as well as the instructional conditions "opportunity to learn", time on task", "homework" and "monitoring at classroom level", were carried out by Witziers and Bosker and published in Scheerens & Bosker, 1997, Ch. 6. The number of studies that were used for these meta-analyses varied per variable, ranging form 14 to 38 studies.

The results in this summary of reviews and meta-analyses indicate that resource-input factors on average have a negligible effect, school factors have a small effect, while instructional have an average to large effect. The conclusion concerning resource -input factors should probably be modified and "nuanced" somewhat, given the results of more recent studies referred to in the above, e.g. the results of the STAR-experiment concerning class-size reduction. There is an interesting difference between the relatively small effect size for the school level variables reported in the meta-analysis and the degree of certainty and consensus on the relevance of these factors in the more qualitative research reviews. It should be noted that the three blocks of variables depend on types of studies using different research methods. Education production function studies depend on statistics and administrative data from schools or higher administrative units, such as districts or states. School effectiveness studies focussing at school level factors are generally carried out as field studies and surveys, whereas studies on instructional effectiveness are generally used on experimental designs. The negligible to very small effects that were found in the re-analysis of the IEA data-set could be partly attributed tot the somewhat "proxy" and superficial way in which the variables in question were operationalized as questionnaire items. An additional finding from international comparative studies (not shown in the table) is the relative inconsistency of the significance of the school effectiveness correlates across countries, also see Scheerens, Vermeulen and Pelgrum, 1989 and Postlethwaite and Ross, 1992.

Table 6: *Review of the evidence from qualitative reviews, international studies and research syntheses*

	Qualitative reviews	International analyses	Research syntheses
Resource input variables:			
Pupil-teacher ratio		-0.03	0.02
Teacher training		0.00	-0.03
Teacher experience			0.04
Teachers' salaries			-0.07
Expenditure per pupil			0.20
School organizational factors:			
Productive climate culture	+		
Achievement pressure for basic subjects	+	0.02	0.14
Educational leadership	+	0.04	0.05
Monitoring/evaluation	+	0.00	0.15
Cooperation/consensus	+	-0.02	0.03
Parental involvement	+	0.08	0.13
Staff development	+		
High expectations	+	0.20	
Orderly climate	+	0.04	0.11
Instructional conditions:			
Opportunity to learn	+	0.15	0.09
Time on task/homework	+	0.00/-0.01 (n.s.)	0.19/0.06
Monitoring at classroom level	+	-0.01 (n.s.)	0.11 (n.s.)
Aspects of structured teaching:			
-cooperative learning			0.27
-feedback			0.48
-reinforcement			0.58
Differentiation/adaptive instruction			0.22

Evidence from developing countries

In this part of the chapter the evidence about effectiveness enhancing conditions of schooling in developing countries will be reviewed. The review sets out by referring to earlier review articles, particularly those by Hanushek (1995) and by Fuller and Clarke (1994), which in

itself incorporates results of reviews by Fuller (1987), Lockheed & Hanushek (1988), and Lockheed & Verspoor (1991). Next a schematic description of 13 studies conducted after 1993 is provided. Conclusions are drawn about the state of the art of educational effectiveness research in developing countries, in terms of predominance of the type of factors that are studied, outcome comparison with results from industrialized countries, relevant research innovations and implications for policy and practice applications.

Production function studies in developing countries

Hanushek (1995) provides the following tabulation of the effects of resources in 69 studies in developing countries (see Table 7).

Table 7: *Summary of ninety-six studies on the estimated effects of resources on education in developing countries, cited from Hanushek, 1995*

		Statistically significant		Statistically insignificant
Input	Number of studies	Positive	Negative	
Teacher-pupil ratio	30	8	8	14
Teacher's education	63	35	2	26
Teacher's experience	46	16	2	28
Teacher's salary	13	4	2	7
Expenditure per pupil	12	6	0	6
Facilities	34	22	3	9

When the number of positive significant associations are expressed in percentages the comparison depicted in table 8 with the results shown in Table 1 concerning studies in industrialized countries can be made in a more straightforward way.

Table 8: *Percentages of studies with positive significant associations of resource input variables and achievement for industrialized as compared to developing countries (sources: Hanushek, 1995, 1997)*

Input	**Industrialized countries**	**Developing countries**
	% sign. positive associations	% sign. Positive associations
Teacher/pupil ratio	15%	27%
Teacher's education	9%	55%
Teacher's experience	29%	35%
Teacher's salary	20%	30%
Per pupil expenditure	27%	50%

The relevance of facilities in education in developing countries, not shown in the comparison, amounts to no less than 70 when expressed as the percentage of significant positive studies. The larger impact of these resource input factors in developing countries can be attributed to larger variance in the independent as in the dependent variables. Both human and material resources in education in industrialized countries are distributed in a relatively homogeneous way among schools, in other words: schools do not differ that much on these variables. Regarding the outcome variables (e.g. educational achievement) Riddell (1997) has shown that schools in developing countries vary on average 40% (raw scores) and 30% (scores adjusted for intake variables). This is a considerably larger variation than is usually found in industrialized countries; where values of 10% to 15% between school variance on adjusted outcomes are more common (cf. Bosker & Scheerens, 1999).

The positive outcomes of production function studies in developing countries make intuitive sense (if basic resources and facilities are not present this will obviously be detrimental to the educational endeavor as a whole). At the same time the outcomes give rise to interesting interpretations when they are brought to bear on the theoretical principles of micro-economic theory. Jimenez & Paquea (1996), for example, present findings that support the thesis that local involvement in school finance stimulate both achievement orientation as economy in spending. Pritchett and Filmer (1997) point at the political advantages of spending on human resources (diminishing class size in particular) as compared to spending on instructional materials, despite the much larger efficiency of the latter approach, while Picciotto (1996)

criticizes the narrow set of educational performance criteria that is used in most education production function research and states that "program design must be informed by assessments of overall educational performance against societal objectives; by evaluations of the relevance of the objectives themselves and by judicious design of institutions to deliver the needed services" (ibid, 5). Micro-economic theory has interesting conjectures with respect to control mechanisms in education as well; where the argument is that bureaucratic control measures are expensive and faulty and community involvement and "direct democracy" would present a better alternative. Particularly when studies are becoming more theory-driven and cost-benefit analyses are more frequently included, production function research is to be considered as a viable approach to school effectiveness studies in developing countries.

reviews of school effectiveness research in developing countries

The results of the review study by Fuller and Clarke (1994) are summarized in Table 9.

Table 9: *School input and process variables that showed significant positive associations with achievement in at least 50% of the studies in developing countries, analyzed by Fuller and Clarke, 1994*[*)]

	Number of significant effects divided by the number of analyses	
SCHOOL/TEACHER FACTOR	Primary Schools	Secondary Schools
School spending		
Expenditure per pupil	3/6	3/5
Total school expenditure	2/5	-
Specific school inputs		
Average class size	9/26	2/22
School size	7/8	1/5
Availability of textbooks	19/26	7/13
Supplementary readers	1/1	2/2
Exercise books	3/3	-
Teaching guides	0/1	-
Desks	4/7	0/1
INSTRUCTIONAL MEDIA	3/3	-
Quality of facilities	6/8	1/1
School library	16/18	¾
Science laboratories	5/12	1/1
Child nutrition and feeding	7/8	1/1

Teacher attributes		
Total years of schooling	9/18	5/8
Earlier measured achievement	1/1	1/1
Tertiary or teacher college	21/37	8/14
In-service teacher training	8/13	¾
Teacher subject knowledge	4/4	-
Teacher gender (female)	½	2/4
Teacher experience	13/23	1/12
Teacher salary level	4/11	2/11
Teacher social class	7/10	-
Classroom pedagogy and organization		
INSTRUCTIONAL TIME	15/17	12/16
Frequent monitoring of pupil performance	¾	0/1
Class preparation time	5/8	½
Frequency homework	9/11	2/2
Teacher efficacy	1/1	0/1
Cooperative learning task student	-	3/3
School Management		
School cluster membership	2/2	-
Principal's staff assessment	¾	0/1
Principal's training level	¾	½
School inspection visits	2/3	0/1
Tracking or pupil segregation	1/1	-

*) Source: Fuller & Clarke, 1994.

The review considered about 100 studies and drew upon earlier reviews by Fuller (1987), Lockheed & Hanushek, 1988, Lockheed & Verspoor and an analysis of 43 studies in the period 1988-1992 conducted by the authors themselves. Only studies that controlled achievement for students' family background were included; and only significant associations at the 5% level were reported. What table 9 indicates is, first of all, that there were more studies about primary schools than about secondary schools. Also, financial, material and human resource input variables were investigated more frequently than school and classroom process variables, with the exception of instructional time. This predominance of relatively easily assessable input characteristics is also evident from Table 10 where the number of times a particular variable was included in a total of 43 studies is indicated.

Table 10: *The number of times out of a total of 43 studies conducted between 1988 and 1992 (primary and secondary schools taken together) a particular type of school input or process variable was investigated. Source: Fuller & Clarke, 1994*

Enrolments/staff	
SCHOOL SIZE	
Class size	6
Teacher variables	
Teacher training	24
Teacher salaries	3
Teacher experience	9
Teacher preparation	1
Teacher efficacy	1
Teacher gender	5
Inservice training	7
Instruction	
Instructional time	13
Homework	3
Specific pedagogy	12
TESTING OF PUPILS	5
School organization	
Public/Private	4
Tracking	1
Headmaster supervision	3
Equipment and facilities	
Library facilities	3
General facilities and equipment	15

On the basis of their review of significant positive effects Fuller and Clarke (ibid) conclude that rather consistent school effects can be found in three major areas: *availability of textbooks and supplementary reading material, teacher qualities* (e.g. teachers' own knowledge of the subject and their verbal proficiencies) and *instructional time and work*

demands placed on students. Policy relevant factors that showed inconsistent or lack of effects appeared to be class size and teacher salaries.

The findings summarized in tables 10 and 11 once more underline the predominance of production function type of effectiveness studies in developing countries. Riddell (1997), in a more methodologically oriented review, observes that a "third wave" of school effectiveness research in developing countries is "in danger of being lost without ever having been explored". By this third wave she refers to, what I have described as "integrated school effectiveness studies", comprising resource inputs, organizational factors and instructional characteristics, in which multi-level modeling is a vital methodological requirement. An interesting set of suggestions that Fuller & Clarke develop in their interpretation of the research evidence, is to pay more attention to cultural contingencies when studying school effectiveness in developing countries. Such contingencies might help in explaining why school and classroom level variables "work" in one country but not in the next. They distinguish four broad categories of cultural conditions:

- the local level of family demand for schooling;

- the school organization's capacity to respond to family demand "while offering forms of knowledge that are foreign to the community's indigenous knowledge" (Fuller & Clarke, 1994, p. 136);

- the teacher's capacity and preference for mobilizing instructional tools;

- the degree of consonance between the teacher's pedagogical behavior and local norms regarding adult authority, didactic instruction and social participation within the school (ibid, p. 136).

These ideas, as well as the appeal to overcome other weaknesses of school effectiveness studies (lack of cost benefit analyses, shortage of longitudinally designed studies) have demanding implications for the design of studies. According to Riddell (1997) Fuller and Clarke fail to present clear research alternatives.

From a review of 12 more recent effectiveness studies carried out in developing countries (Scheerens, 1999) reconfirmed the predominance of the production function approach with a restatement of the importance of equipment, particularly textbooks and the human resource factor (teacher training). According to the author instructional and pedagogical theory appeared to be practically missing as a source of inspiration for educational effectiveness studies in developing countries. In the four studies that did look into some school organizational and instructional variables , the impact of these variables was relatively low. This (limited) review of 12 studies confirms the results of an earlier review by Anderson, Ryan and Shapiro (1989) who stated that "variations in teaching practice in developing

countries are only rarely found to be associated with variations in students learning". Cultural contingencies, as referred to by Fuller and Clarke, or lack of variation in teaching practices in some developing countries could be offered as hypothetical explanations for these outcomes.

Summary and conclusion

In this chapter five strands of educational effectiveness research were discussed. The general conclusion, when reviewing the bulk of the research, was that in developed countries the impact of resource-input factors is fairly small. This outcome was interpreted against the background of relatively small variation in these variables in developed countries. On the basis of recent studies, human resource inputs, particularly teacher qualifications, deserve reconsideration, however. In developing countries the significance of the impact of resource input factors was established in a larger proportion of studies. Several reviewers have pointed at the larger between school differences in developing countries (Bosker & Witziers, 1996, Riddell, 1997), which could explain the differences between developed and developing countries in these research outcomes.

Compensatory programs, school improvement projects and studies of unusually effective schools in developed countries have concentrated on a similar set of relevant school-organizational variables. Reviewers agree on the relevance of factors like: achievement oriented school policy, educational leadership, consensus and cooperation among staff, opportunities for professional development of staff and parental involvement. When subjected to statistical meta-analysis, the impact of these school-organizational factors is relatively small to "medium". In developing countries these factors have been studied infrequently; what results are available show insubstantial impact.

At classroom level instructional and teacher effectiveness studies have indicated medium to large effects of variables like: time on task, content covered or "opportunity to learn", and aspects of structured teaching like; frequent monitoring of students' progress, feedback , reinforcement and cooperative learning. A limitation of these research outcomes is that they have not addressed other than subject-matter based learning objectives in traditional school subjects. On the other hand such learning objectives are likely to remain relevant and these outcomes, which support a behaviouristic interpretation, are sufficiently robust to be considered vis a vis constructivist perspectives on learning and instruction. Again, results depend mostly on studies in developing countries. From the limited number of studies in developing countries that was considered (for a more detailed review see part II if this chapter) no substantive impact of instructional factors was apparent. More detailed and in depth studies of instructional variables in the context of developing countries, also in relationship to cultural background factors, as suggested by Fuller and Clarke, 1994, are considered as quite relevant for future research.

In the course of this chapter quite a few limitations of the research findings have been pointed out, also with respect to the interpretation and use of these findings in developing countries. The question of the robustness of the knowledge base on school effectiveness should, once again, be considered.

What is to be noted, first of all, is that in developed countries the margins to which schools can make a difference appear to be relatively small when expressed in the usual social scientific criteria for effect-sizes. The "net" between school variance, i.e. the proportion of variance in achievement at the student level that can be attributed to attending a particular school, after adjustment for relevant background variables, is estimated as low as 4% (Bosker & Witziers, 1996). When interpreted in a more "practical" way, for example by comparing the 10% most effective schools to the 10% least effective schools, for a country like the Netherlands, would make for a difference of one or two levels of the hierarchically categorized secondary school-system. Other authors have expressed this difference in terms of one grade-level (Purkey & Smith, 1983). It should also be noted that this societal effect would be there for all the pupils in these 10% higher or lower scoring schools. The next question is the degree to which the net between school variance in pupils'achievement is attributable to the malleable conditions of schooling that are considered as the "independent" variables. In a typical "integrated" school effectiveness study, which contains school level and classroom level variables, as the study by Brandsma, 1993, the relevant proportion was about 60%. An important alternative source of variance being the "contextual" effect of e.g. the average initial aptitude of the students. Within the small margins of the variance that lies between schools in developed countries, this appears to be a fair support for the variables that have been proposed as hypothetical effectiveness enhancing conditions.

In developing countries research appears to support the common sense notion that provision of basic resources, particularly among the most deprived schools, makes most of the difference. In this context the challenge for the future lies in more frequent and in-depth study of instructional conditions.

A final observation regards the larger impact of factors closer to the actual teaching and learning process as compared to more 'distal' factors like school organizational and school environmental conditions. From the perspective of national policy-making and planning these results should be weighted against the efficiency of bringing about changes at a higher level in the system (which contains fewer units). If there is evidence for a positive, although small, significant impact of a particular style of school leadership, "instructional" or "educational" leadership as this research literature shows, a training course for head teachers could be more cost-effective than training all the teachers in the country.

PART II REVIEW OF INSTRUCTIONAL EFFECTIVENESS RESEARCH

This part of the chapter has sections on early reviews, modelling instructional effectiveness, quantitative evidence on effect sizes, review of studies in developing countries and a review of more recent studies. In the final section the issue of the influence of national policy on stimulating teacher and teaching effectiveness is discussed

Early Reviews of research on teaching

In the sixties and seventies the effectiveness of certain personal characteristics of teachers was particularly studied. Medley & Mitzel (1963); Rosenshine & Furst (1973) and Gage (1965) are among those who reviewed the research findings. From these it emerged that there was hardly any consistency found between personal characteristics of the teacher like being warm hearted or inflexible on the one hand, and pupil achievement on the other. When studying teaching styles (Davies, 1972), the behavioural repertoire of teachers was generally looked at more than the deeply-rooted aspects of their personality. Within the framework of "research on teaching" there followed a period in which much attention was given to observing teacher behaviour during lessons. The results of these observations, however, in as far as they were related to pupil achievement, seldom revealed a link with pupil performance (see Lortie, 1973, for instance). In a following phase more explicit attention was given to the relation between observed teacher behaviour and pupil achievement. This research is identified in the literature as "process-product studies". Lowyck, quoted by Weeda (1986, p. 68), summarises variables which emerged "strongly" in the various studies:

1. *Clarity:* clear presentation adapted to suit the cognitive level of pupils.
2. *Flexibility:* varying teaching behaviour and teaching aids, organising different activities etc.
3. *Enthusiasm:* expressed in verbal and non-verbal behaviour of the teacher.
4. *Task related and/or businesslike behaviour:* directing the pupils to complete tasks, duties, exercises etc. in a businesslike manner.
5. *Criticism*: much negative criticism has a negative effect on pupil achievement.
6. *Indirect activity*: taking up ideas, accepting pupils' feelings and stimulating self-activity.
7. *Providing the pupils with an opportunity to learn criterion material* - that is to say, a clear correspondence between what is taught in class and what is tested in examinations and assessments.
8. Making use of *stimulating* comments: directing the thinking of pupils to the question, summarising a discussion, indicating the beginning or end of a lesson, emphasising certain features of the course material.
9. *Varying the level* of both cognitive questions and cognitive interaction.

Weeda (1986, p. 69) observes that in the study from which these nine teaching characteristics were drawn, there was much criticism regarding methodology/technique. He divides the later research studies focused at instructional effectiveness into two areas:

- pedagogic studies aimed at tracing certain environmental factors and teaching behaviour that can influence levels of performance of certain groups of pupils;
- instructional psychology research aimed at establishing the interaction between teaching variables and pupil characteristics; the so-called *aptitude-treatment-interaction* studies.

A central factor within the first area is that of effective teaching time. The theoretical starting points of this can be traced back to Carroll's teaching-learning model (Carroll, 1963). Chief aspects of this model are:
- actual net learning time which is seen as a result of:
  perseverance and opportunity to learn;
- necessary net learning time as a result of:
  pupil aptitude, quality of education and pupil ability to understand instruction.

The mastery learning model formulated by Bloom in 1976 was largely inspired from Carroll's model.
The findings of the aptitude-treatment-interaction studies were generally judged to be disappointing. There were scarcely any interactions discovered, which was later confirmed by a replication study. De Klerk (1985) regarded the fact that the ATI had failed to reveal any simple interaction between pupil characteristics and instruction method as a challenge to do more refined empirical research on more complex interaction patterns.

Stallings (1985) summarised research literature on effective instruction - in as far as it was concerned with primary education - under the headings: effective net learning time, class organisation and management, instruction, assessment and teacher expectations.
When studying net learning time it emerged that simply making the school day longer did not necessarily lead to better levels of performance. More important, ultimately, is how effectively time is spent. Stallings and Mohlman (1981) established that effective teachers spent 15% of the school day on organisation and management; 50% on interactive teaching and 35% on monitoring pupils' work. Aids for an effective use of instruction time include all types of lesson planning. Under the classification *class organisation and management* Stallings discusses streaming and maintaining order. Studies on streaming or working with ability groups as compared to whole class instruction indicate that this type of teaching works more positively with the more gifted pupils and that with less able groups - taking the average result of the large numbers of surveys - hardly any effect was found (also according to Kulik & Kulik, 1982, Van Laarhoven & De Vries, 1987, Reezigt, 1993 and Slavin, 1987). Moreover, from various types of studies it emerges that in classes where there is disruptive behaviour, pupil performance is lower: disruption, naturally enough, is at the cost of effective learning time.
The question what makes good teaching should be looked at on different levels. For direct question-and-answer type knowledge other teaching strategies are called for than for problem-solving and acquiring insight. For learning tasks which greatly depend on memory, a highly ordered and consistent approach is the most effective. For the acquiring of insight too a clear presentation of the information offered is important as are questions to check whether pupils have actually absorbed a specific insight. With regard to problem-solving, some empirical

support is available which shows that it is desirable that pupils take much initiative themselves. Collins & Stevens (1982) mention five teaching strategies to support learning in the sense of problem-solving: a) a systematic variation of examples; b) counter examples; c) entrapment strategies; d) hypothesis identification strategies; e) hypothesis evaluation strategies.

From studies on *teacher assessments and expectations* of pupils it seems that self-fulfilling prophecies can occur. If a teacher has once formed negative expectations of certain pupils (s)he is likely to give them less attention and expose them less to more difficult and challenging tasks. Obviously this is even more of a disadvantage if the initial assessment was a wrong one. Thus it is imperative that teachers should try and avoid negative stereotyping of pupils (Van der Hoeven-Van Doornum, 1990).

In a review of literature on effective teaching at *secondary school level* Doyle (1985) deals broadly with the same categories as Stalling's, namely "*time on task*" and "*quality of instruction*". Because in secondary education the total teaching spectrum from which a choice must be made is far greater than in primary education, the variable "opportunity to learn" is associated here with the concept of effective net learning time. "Opportunity to learn" is generally understood in the sense of offering pupils a range of subjects and tasks that cover educational goals. In educational research, opportunity to learn concentrates on the extent to which classroom exercises correspond with the content of the tests for monitoring performance.

As far as the quality of instruction is concerned, there is a stronger emphasis in secondary education on learning higher cognitive processes like insight, flexibly adopting knowledge and problem-solving. Doyle considers the effectiveness of direct teaching, which he defines as follows:

1. Teaching goals are clearly formulated.
2. The course material to be followed is carefully split into learning tasks and placed in sequence.
3. The teacher explains clearly what the pupils must learn.
4. The teacher regularly asks questions to gauge what progress pupils are making and whether they have understood.
5. Pupils have ample time to practice what has been taught, with much use being made of "prompts" and feedback.
6. Skills are taught until mastery of them is automatic.
7. The teacher regularly tests the pupils and calls on the pupils to be accountable for their work.

The question whether this type of highly structured teaching works equally well for acquiring complicated cognitive processes in secondary education can be answered in the affirmative (according to Brophy & Good, 1986, p. 367). However, progress through the subject matter can be taken with larger steps, testing need not be so frequent and there should be space left for applying problem-solving strategies flexibly. Doyle also emphasises the importance of varying the learning tasks and creating intellectually challenging learning situations. For the latter an evaluative climate in the classroom, whereby daring to take risks even with a complicated task is encouraged, is a good means. In addition, Doyle deals with the effect of certain ways of working and grouping, including individual teaching and working together in small groups. Bangert, Kulik & Kulik's meta-analysis (1983) revealed that individualized teaching in secondary

education hardly led to higher achievement and had no influence whatsoever on factors like the self-esteem and attitudes of pupils.
Evaluation studies on special programmes to stimulate working in small groups reveal that some of these have a positive effect on lower attaining pupils. Generally speaking, from other reviews of research on the effects of cooperative learning it appears that there is no conclusive empirical evidence to support the positive influence of this type of work on performance. Vedder (1985) explained the lack of an unequivocal positive influence of group work by the possible fact that due to the way pupils work together there is insufficient cognitive stimulation present.

Modelling Instructional effectiveness

The Carroll model (Carroll, 1963) is usually considered as the starting point of modelling instructional effectiveness. It consists of five classes of variables that are expected to explain variations in educational achievement. All classes of variables are related to the time required to achieve a particular learning task. The first three factors are directly expressed in terms of amounts of time, while the two remaining factors are expected to have direct consequences for the amount of time that a student actually needs to achieve a certain learning task. The five classes of variables are:

- *aptitude*; variables that determine the amount of time a student needs in order to learn a given task under optimal conditions of instruction and student motivation;
- *opportunity to learn*; the amount of time allowed for learning;
- *perseverance*; the amount of time a student is willing to spend on learning the task or unit of instruction.
- *quality of instruction;* when the quality of instruction is sub-optimal, the time needed for learning is increased;
- *ability to understand instruction*, e.g. language comprehension, the learners' ability to figure out independently what the learning task is and how to go about learning it (Carroll, 1963, 1989).

Numerous research studies and meta-analyses have confirmed the validity of the Carroll model (see chapter 5). The Carroll model has also been the basis for Bloom's concept of mastery learning (Bloom, 1968) and is also related to "direct instruction", as described by Rosenshine (1983).
Characteristics of mastery learning are:
1) Clearly defined educational objectives.
2) Small discrete units of study.
3) Demonstrated competence before progress to later hierarchically related units.
4) Remedial activities keyed to student deficiencies.
5) Criterion-referenced rather than norm-referenced tests (Block & Burns, 1970).
Direct instruction also emphasizes structuring the learning task, frequent monitoring and feedback and high levels of mastery (success rates of 90 to 100% for initial tasks) in order to boost the self-confidence of the students.

The one factor in the original Carroll model that needed further elaboration was "quality of instruction". As Carroll pointed out himself in a 25-year retrospective of his model, the original formulation was not very specific about the characteristic of high-quality instruction. "But it mentions that learners must be clearly told what they are to learn, that they must be put into adequate contact with learning materials, and that steps in learning must be carefully planned and ordered" (Carroll, 1989, p. 26).
The cited characteristics of mastery learning and direct instruction are to be seen as a further operationalization of this particular factor, which is of course one of the key factors (next to providing optimal learning time) for a prescriptive use of the model. It should be noted that Carroll's reference to students who must be put into adequate contact with learning materials, developed into a concept of "opportunity to learn" different from his own. In Carroll's original formulation, opportunity to learn is identical to allocated learning time, while now opportunity to learn is mostly defined in terms of the correspondence between learning tasks and the desired outcomes. Synonyms for this more common interpretation of opportunity to learn are: "content covered" or "curriculum alignment" (Berliner, 1985, p. 128). In more formal mathematical elaborations the variable "prior learning" has an important place (Aldridge, 1983; Johnston and Aldridge, 1985).
The factor allocated learning time has been further specified in later conceptual and empirical work. Karweit and Slavin (1982), for instance, divide *allocated learning time* (the clock time scheduled for a particular class) into *procedural time* (time spent on keeping order, for instance) and *instructional time* (subject matter related instruction) and *time on task* (the proportion of instructional time during which behaviour appropriate to the task at hand took place).
Ability to understand instruction can be seen as the basis for further elaboration in the direction of learning to learn, meta-cognition, etc. The comprehensiveness of the Carroll model is shown by this potential to unite two schools of instructional psychology, the behaviouristically inclined structured teaching approaches and the cognitivist school (cf. Bruner, 1966; De Corte & Lowyck, 1983).

As stated in the above, starting from the initial Carroll model, an important development has been to further "fill in" the black-box of "quality of instruction". Making use of the principles of mastery learning and direct instruction, Creemers (1994) has proposed a more elaborate model in which three main aspects of "quality of instruction" are distinguished: curriculum, grouping procedures and teacher behaviour. Each of these components contains a set of effectiveness-enhancing conditions, which is roughly similar across the three components. Creemers calls this the consistency principle: "..... the same characteristics of effective teaching should be apparent in the different components. It is even more important that the actual goals, structuring, and evaluation in curricular materials, grouping procedures, and teacher behaviour are in the same line. ..... In this way a synergetic effect can be achieved." (ibid, p. 11).

Structuring and the cybernetic cycle of evaluation, feedback and corrective action can be seen as the basic factors behind instructional quality in each of the three domains.

Another more recent development in modelling instructional effectiveness is the emerging new paradigm inspired by constructivism. Constructivism claims that reality is more in the mind of the knower, but does not go as far as denying external reality altogether (solipsism), however some radical constructivists do come very close to a position of complete denial. The image of *student learning* that goes with constructivism underlines the active role of the learner. Students are to be confronted with "contextual" real-world environments, or "rich" artificial environments simulated by means of interactive media. Learning is described as self-regulated with lots of opportunity for discovery and students' own interpretation of events.
Learning strategies, learning to learn and reflecting on these learning strategies (meta-cognition) are as important as mastering content. Different ways in finding a solution are as important as the actual solution itself. Terms like "active learning" (Cohen, 1988), "situated cognition" (Resnick, 1987) and "cognitive apprenticeship" (Collins et al., 1989) are used to describe student learning.

The other side of the constructivist coin are approaches to *teaching* and *instructional technology* that enable students "to construct their own meaningful and conceptually functional representations of the external world" (Duffy and Jonassen, 1992, p. 11). The teacher becomes more of a coach, who assists students in "criss-crossing the landscape of contexts", looking at the concept from a different point of view each time the context is revisited (Spiro et al., 1992, p. 8). Cohen (1988) uses the term "adventurous teaching" for this approach.
There is less emphasis on structuring goals, learning tasks and plans in advance; goals are supposed to emerge when situated learning takes place and plans are not so much supposed to be submitted to the learner as constructed in response to situational demands and opportunities.
Learning situations must be such that students are invited to engage in sustained exploration (real-life contents, or simulated environments). Some authors writing from this perspective state that "transfer" is the most distinguishing feature (Tobias, 1991), whereas others mention argument, discussion and debate to arrive at "socially constructed meaning" (see Cunningham, 1991).
The role of assessment and the evaluation of students' progress is hotly debated. Radical constructivists take the position that performance on an actual learning task is the only legitimate way of assessment, since distinct "external" evaluation procedures could not do justice to the specific meaning of a particular learning experience for the student.
Others (e.g. Jonassen, 1992) come to the conclusion that assessment procedures from a constructivist perspective should merely be different: goal-free, rather than fixed on particular objectives, formative rather than summative, and oriented to assessing learning processes rather than mastery of subject matter. Appraisals of samples of products, portfolios and panels of reviewers that examine authentic tasks are also mentioned as acceptable procedures.

In Table 1 some of the major distinguishing features of learning and instruction according to the constructivist position are contrasted with characteristics of more traditional instructional models like direct instruction and mastery learning.

Table 1: *Comparison of traditional and constructivist instructional models; source Scheerens, 1995*

Traditional instruction	*Instruction inspired by constructivism*
emphasis on basic skills	bias towards higher order skills
subject matter orientation	emphasis on learning process
structured approach • prespecified objectives • small steps • frequent questioning/feedback • reinforcement through high % of mastery	discovery-learning "rich" learning environment • intrinsic motivation • challenging problems
abstract-generalizable knowledge	situation-specific knowledge, learning from cases
standardized achievement tests	assessment; less circumscribed alternative procedures

Bipolar comparisons such as the one in Table 1 run the risk of over - simplification and polarization whilst also constructing "straw men". It should be emphasized that less extreme constructivist views can be very well reconciled with more "objectivist" approaches (cf. Merrill, 1991). Also, more eclectic approaches are feasible, as can be seen when more teacher-controlled and learner-controlled instructional situations are used alternately (cf. Boekaerts & Simons, 1993).

Creemers (1996) considers the changed perspective on the role of the student as the essential difference between the newer, constructivist views on learning and instruction, and the older models: rather passive in the models originating from the Carroll model and active, picturing a student who conducts knowledge and skills through working with context, in the newer models. Brophy (1996) also points at a way to integrate the established principles of structured classroom management and self-regulated learning strategies. Elements of effective classroom management such as "preparation of the classroom as a physical environment suited to the nature of the planned academic activities, development and implementation of a workable set of housekeeping procedures and conduct rules, maintenance of student attention and participation in group lessons and activities, and monitoring of the quality of the students' engagement in assignments and of the progress they are making toward intended outcomes", are equally relevant when instruction is seen as helping students to become more autonomous and self-regulated learners (ibid, p. 3, 4).

When it comes to implementing the new instructional principles, Brophy points at a "guided", gradual approach where learning goals and expectations are clearly articulated, and students are helped by means of modelling and providing cues. He also stresses the fact that, initially, students may need a great deal of explanation, modelling and cuing of self-regulated learning strategies. As they develop expertise, this "scaffolding" can be reduced.

The quantitative evidence on instructional effectiveness

In the nineteen eighties, several influential research syntheses were carried out by Walberg (1984) and Fraser et al. (1987). The teaching conditions for which Walberg found the highest effects were:
- reinforcement (reward and punishment);
- special programmes for the educationally gifted;
- structured learning of reading;
- cues and feedback;
- mastery learning of physics; and
- working together in small groups.

Fraser et al. (1987) even provided a synthesis of 134 meta-analysis which together comprised 7827 individual studies. Part of their findings are summarised in Table 2.

Table 2: *Effects of teaching and pupil characteristics on performance tests, based on Fraser et al., 1987.*

Factor	*Result* (in correlations)
School characteristics	.12
Social background characteristics of pupil	.19
Teacher characteristics	.21
Teaching characteristics	.22
Pupil characteristics	.24
Instruction method	.14
Learning strategies	.28

Specific variables, included in the main categories in Table 2 which correlate highly with achievement, are: quality of teaching, r = .47; amount of instruction, r = .38; cognitive background characteristics, r = .49 and feedback, r = .30.
A remarkable conclusion that Walberg attaches to the research syntheses he carried out is the statement that the findings apply for all types of schools and all types of pupils. Walberg expresses this in the saying "What's good for the goose is good for the gander". He does add that this especially applies to the more powerful factors (that is the factors that correlate the highest

with levels of performance). When we look at these powerful factors, it seems that highly structured learning or direct teaching, which emphasises testing and feedback, again emerges as the most effective teaching form. Yet, in Walberg's research syntheses there are also forms of individual teaching and teaching adapted to fit the specific needs of pupils as well as working together in small groups that come quite strongly to the fore. He even comes up for "open teaching" in which cooperation, critical thinking, self-confidence and a positive attitude are important objectives. His own and other meta-analyses reveal that open teaching has no adverse consequences for cognitive achievement, while there is a positive influence on creativity, social behaviour and independence. In the meta-meta analysis of Fraser et al. individualising emerged as a less powerful factor ($r = .07$).

No matter how impressive the huge data files may be upon which the research-syntheses are based, there are, nevertheless, limitations attached to the findings. Every time simple correlations are presented whereby it cannot be ruled out that a particular correlation is carried for the most part by a third variable, which in these simple analyses cannot be made visible. This problem exists partly because it can be assumed that many of the individual effectiveness predictors are correlated among themselves. And where this problem applies to the general analyses it can by no means be ruled out that this is also the case with many of the individual studies upon which the syntheses are based.

Finally, with regard to this survey of instructional effectiveness it must again be pointed out that within the scope of this chapter only a broad summing up of the most important research findings on instructional effectiveness is possible. Even if the conclusion is that a few prominent characteristics of effective teaching can be distinguished - the amount of instruction and a structured approach - that apply to any given teaching situation, nevertheless, it should certainly not be forgotten that with a less general treatment all types of nuances exist that are linked to differences in subjects taught, pupil characteristics, school type and educational goals. For a review in which these nuances are well expressed reference is made to Brophy & Good, 1986.

In a more recent synthesis of meta-analyses, and reviews (maybe the term *mega*-analysis would be appropriate for this work) Wang, Haertel and Walberg summarize the current knowledge with respect to the influence of educational, psychological and social factors on learning (Wang, Haertel & Walberg, 1993). Although the evidence they present is comprehensive in the sense that school context factors and school level factors are included, the largest amount of studies concern "design and delivery of curriculum and instruction" (36%), "student characteristics" (24%) and "classroom practices" (18%).

A first main outcome of the Wang et al. research synthesis is a rank-ordering of the relative importance of "distal" versus "proximal" factors in influencing achievement. Distal factors are less directly associated with the primary process of learning and instruction, examples are: "state and district governance and organization" and "school demographics, culture, climate, policies and practices". Student characteristics and classroom practices are considered as proximal factors, close to the instructional process. The results of the syntheses show that the more proximal factors have a stronger positive association with educational achievement, as compared

to more distal factors. The rank-ordering Wang et al. present is as follows (ordered from high to low):

- student characteristics
- classroom practices
- home and community educational contexts
- design and delivery of curriculum and instruction
- school demographics, culture, climate, policies and practices
- state and district governance and organization

Leaving aside student characteristics, more specific factors that have the strongest association with achievement are: "classroom management" and "student and teacher social interaction" (both aspects of the more general factor "classroom processes") and "home environment" (aspects of the more general factor "home and community educational contexts"). An illustrative variable within the "classroom management category" is "group alerting" (teacher uses questioning/recitation strategies that maintain active participation by all students). Other variables that are relatively influential within the classroom practice factor are: classroom climate, classroom assessment, quantity of instruction (e.g. time on task) and "student and teacher interaction" (e.g. "students respond positively to questions from other students and from the teacher").

In their interpretation of these effective classroom practices the authors emphasize the following points:

- *academic* student teacher interactions should "make students aware of subject-specific knowledge structures", for example by an appropriate use of questioning by the teacher;
- *social* teacher and student interactions should dissuade students from disruptive behaviour and "establish a classroom atmosphere conducive to learning"; besides the use of praise and corrective feedback is mentioned.

In summary, it appears that, as far as classroom instruction is concerned, this recent meta-analysis enforces the importance of general instruction approaches that are quite structured, like mastery learning and direct instruction. At the same time, the interpretation the authors give to "academic teacher and student interactions", namely in terms of laying bare "knowledge structures" and the importance of "meta-cognition" as a student background factor, provide support to an emerging, more cognitivist/constructivist view on learning and instruction.

In comparison to the sizes of the effect of resource input variables, the effects of instructional conditions, reported in meta-analyses, are generally larger (correlations in the order of .20). In most instances, however, the associations reported reflect relationships unadjusted for student background characteristics.

One of the strong points of these mega-meta-analyses, is that information with a wide scope is provided, while using global categories of effectiveness enhancing conditions. At the same time, however, this characteristic may also be taken as one of the limitations of this approach. The results do lack a certain degree of specificity. Another point of criticism is that the varying research quality of the studies used as a basis for meta-analysis is not always taken into account.

In order to overcome some of these disadvantages of meta-analysis Slavin (1996) has introduced an approach which he calls "best-evidence synthesis". He describes this approach as a combination of the methods of systematic literature review and meta-analysis. In "best-evidence synthesis" the quantification of study outcomes is combined with the discussion of substantive and methodological issues of narrative reviews (ibid, p. 3). In his book "Education for All" (Slavin, 1996), best-evidence syntheses on several instructional practices are presented: the instructional practices in question are: co-operative learning, ability grouping in elementary and secondary schools and mastery learning.

With respect to *cooperative learning*, Slavin analyzed a total of 99 studies in which different types of cooperative learning were investigated. Of these 99 studies 64% showed a significant advantage of the experimental group over the control group. In only 5% of the study control groups were favoured. The median effect sizes for different types of cooperative learning range from .04 of a standard deviation to .86 of a standard deviation. The mean effect size over all studies was .26 s.d., which can be seen as an educationally significant effect.
Factors that make group work "work", are the use of group rewards based on the individual learning of all group members and the direct teaching of structured methods to students to work together or by teaching them learning strategies closely related to the instructional objectives (ibid, p. 57).

Applying the same method of "best-evidence synthesis" Slavin (1996) has also analyzed ability grouping, by analyzing studies in which a form of ability grouping was compared to a control condition of heterogeneously grouped classes. His conclusion is that ability grouping between classes in order to create groups that are homogeneous in ability level or to use his term "*ability-grouped class assignment*" is generally ineffective (effect sizes are either negative or close to zero, with only a few exceptions of studies reporting small positive effects).
Grouping for reading across grade lines (the so-called Joplin Plan) has a consistently positive effect (median effect size = +.44 s.d.), the same conclusion is drawn with respect to *within-class ability grouping* in mathematics in the upper elementary school (median effect size = +.34 s.d.).
In his explanation as to *why* certain types of grouping (e.g. within classes) work and others (ability-grouped class assignment) do not, Slavin applies three criteria:

1) grouping must reduce heterogeneity in the specific skill being taught;
2) the grouping plan must be flexible enough to correct for misassignments and changes in student performance;
3) "Teachers must actually vary their pace and level of instruction to correspond to students" levels of readiness and learning rates" (ibid, p. 158).

He concludes that ability-grouped class assignment fails the first and probably also the second criterion.
Apart from ability-grouped class assignment being generally ineffective, it is also disadvantageous in the sense that it is likely to cause segregation and institutionalize low teacher expectations.

Relevant explanatory background factors with respect to the question as to why certain types of grouping are effective and others ineffective, considered by Slavin are instructional time, and the use of assessment-based adaptive instruction. The latter characteristic is likely to be conducive to instruction that is closely tailored to students' levels of performance. With respect to instructional time, within class ability grouping will usely require some loss of time for transition of the teacher from one group to the next, and will also imply that groups will have to work a considerable amount of time without direct teacher instruction. Slavin's results indicate that within class ability grouping is more effective than whole class teaching to heterogeneous classes, despite a certain loss of direct instructional time. To suppress the negative impact of loss of instructional time there should not be too many ability groups within the class.
On the basis of these findings Slavin puts forward the following tentative recommendations (ibid, p. 164):

1) Leave students in heterogeneous classes most of the time and regroup by ability only in subjects (reading, mathematics) in which reducing heterogeneity is particularly important.
2) Grouping plans should reduce heterogeneity in the specific skill being taught.
3) Grouping plans should be flexible and allow for easy reassignment.
4) Teachers should actually vary their level and pace of instruction to correspond to student performance level.
5) In within-class ability grouping, the number of groups should be small.

The lack of effect of ability grouping between classes (streaming or tracking) was reproduced in a "best-evidence synthesis" in which studies on secondary schools were analyzed (Slavin, 1996, pp. 167-188). The median effect-size that was found was -.02.

Mastery learning ranked high in the meta-analyses from Walberg (1984) and Fraser et al. (1987) cited in a previous section. Slavin (1996) conducted a best-evidence synthesis and found moderately positive effects on experimenter-made achievement measures (about 1/4 s.d.), closely tied to the objectives taught in the mastery learning classes, but practically no effects when standardized achievement tests were used as the dependent variable.
He concludes that his findings do *not* support the claim that mastery learning is more effective than traditional instruction given equal time and achievement measures "that assess coverage as well as mastery" (ibid, p. 253). The explanation Slavin gives for the fact that his outcomes are much more modest than those of other reviewers and meta-analysts is that he selected only studies that met the criterium of a study duration of at least four weeks, whereas the other reviewers included many short time studies, and studies in which the effect of mastery learning was inflated because of more instruction time - even one-to-one tutoring - in the experimental group.
So, the selection of studies that Slavin used, were higher in ecological validity than many of the more laboratory-type studies included in the other meta-analyses and reviews. In "real life" classroom situations the major principles of mastery learning, particularly the amount of corrective instruction may not be applied optimally ("too little and too late"; ibid, p. 256). The key-element in mastery learning, according to Slavin, is the frequency of testing and feedback.

Other related elements are well-specified educational objectives and basing teaching decisions on the results of these assessments. And finally the "time"-factor plays a central role in the success of mastery-learning approaches (although this factor has sometimes confounded the outcomes of experimental studies).

The basic factors that emerge from the meta-analyses and "best-evidence synthesis" cited in this chapter are:

- time on task;
- closeness of content-covered to assessment instrument;
- a structured approach: specific objectives, frequent assessment and corrective feedback;
- types of "adaptive" instruction that can be managed by teachers (e.g. no more than two within-class ability groups per classroom).

Again, in comparison to the effect sizes noted for resource input factors, the effects of these various instructional conditions are relatively large.

Instructional effectiveness studies in developing countries

On the basis of a review of 12 educational effectiveness studies in developing countries, Scheerens (1999) concluded that school organisational variables and instructional variables were studied in only 4 of these studies. The rest of the studies were economically oriented education production function studies. The author states that given the expectations one might have considering the impact of school organisational and instructional on the basis of research results obtained in OECD countries, a closer look at studies that have investigated these factors is warranted. Several authors, reflecting on the direction educational effectiveness studies should take in developing countries, Lockheed & Longford, 1989; Riddell, 1997, have also referred to these expectations.

The four studies that are reviewed by Scheerens (1999) are those by Glewwe et al. (1995), Fuller et al. (1994), Nyagura & Riddell (1993) and Van der Werf et al. (1999). The study reviews and conclusions are cited literary in the remaining part of this section. studied more cultural and managerial school organizational variables and variables reflecting teaching practices.

Apart from more descriptive characteristics like qualifications of head teachers, Nyagura & Riddell (ibid) studied the following more “substantive” school organizational variables in Zimbabwen secondary schools: teacher stability, time devoted to school-based in-service activities and professional support to teachers through supervision by the head teacher. Data were collected by means of survey methods, yielding information at school, classroom and student leve.. In a preliminary regression analysis all individual variables were related to English and math achievement. Among the more substantive, theoretically interesting class and school level variables the amount of instructional time devoted to mathematics (class

level) and the amount of supervision of teachers had a significant association with achievement. Contrary to expectations the association of the latter variable with achievement was negative, however. Further multi-level modeling pointed out that just one other class-level variable, the amount of supervised study time afforded by the head teacher had a positive and significant effect in mathematics (not in English). Both at school and classroom level textbook availability and teacher training stood out as the most important factors for either subject.

Glewwe et al. (1995) examined, apart from more physical input variables, pedagogical inputs (curriculum, instructional time and teacher quality), pedagogical processes (teaching practices in the classroom) and school organization, climate and control (school autonomy, work-centered environment, community involvement, orderly environment and school type) in their study of primary schools in Jamaica.
Data on school and classroom level variables were collected by means of a school administrator's and a teacher questionnaire. Effect variables were measured by means of the California Achievement Test (CAT) which measures mathematics computation and reading comprehension. This test yielded comparable scores across all grade levels.
Data were analyzed by means of econometric methods. The results pointed out that physical and pedagogical inputs only "played a marginal role in explaining cross-sectional differences in cognitive skills in Jamaica" (ibid, 249).
In the domain of the pedagogical process variables "doing written assignments in class" had a strong *negative* effect on achievement. Testing students had a weakly significant (10% level) effect and time spent in whole-class instruction a weakly significant negative effect on math achievement. In reading intensity of textbook use and the percentage of teacher time spent testing students had positive effects, and time spent doing written assignments had a negative effect.
Among the school organization, climate and control variables there were no significant effects that reached the 5% level. Significant at the 10% level were: discussing curriculum and staffing issues at staff meetings (positive for mathematics); hours of instructional assistance by the principal and average frequency with which teachers help each other (positive for reading).
The authors conclude that overall, variables measuring pedagogical processes are more often significantly related to student achievement than are physical and pedagogical input variables and school organizational variables (ibid, 250). They refer to the high levels that important inputs in Jamaica are at to explain the relatively small impact of such variables.

In the study of Indonesia primary schools Van der Werf et al. (1999) used observation and interview methods to study school organizational and teaching variables. Data were analyzed by means of multi-level modeling (VARCL-program). Student achievement in mathematics, Bahasa Indonesia and Science was measured in grade 6, by means of standardized tests. Out of 27 school and classroom level variables, all selected on the basis of the school and instructional effectiveness literature, 4 variables had a significant (5% level) association with

achievement in the expected (positive) direction for mathematics (time spent on subject, frequent questioning by the teacher, evaluation of teachers and help with homework); for Bahasa Indonesia there were 3 such variables (innovative teaching, observations in classrooms and voluntary work of teachers); with respect to science 3 variables were significant and positive (innovative teaching, evaluation of school quality and availability of student books), however, 4 other variables had a significant effect with the "wrong" sign (time spent to subject, presentation of content, pupils working and evaluation of teachers.
In two cases variables were significantly positive for two out of three subjects (innovative teaching and evaluation of teachers).

In the study of Botswana's junior secondary schools Fuller et al. (1994) observation methods and survey methods were used to study four blocks of school organizational and teaching variables for their effect on language and math. achievement:

A. material conditions and school inputs
B. teacher background, gender and training levels
C. teaching practices and classroom rules
D. teacher effort and pedagogical beliefs

In block C the following pedagogical behaviors were observed: the complexity of instructional tools utilized by teachers, task demands placed on pupils by the teachers, especially the frequency of active reading and writing exercises, the frequency and complexity of questions asked of pupils, the consistency of the teachers "pedagogical technology" and the use of instructional time.
In block D (teacher effort and pedagogical beliefs) pedagogical philosophies, teacher self-perceptions of competence, job satisfaction and level of efficacy were addressed.
Altogether the set of variables in blocks C and D appeared to have "little explanatory power" (ibid, 368). The only variable in these domains that had a significant effect was teachers' average use of open-ended questions, but contrary to expectation, the sign of the association was negative.

How should these results be interpreted? First of all, four studies is obviously too small a number to draw firm conclusions. Nevertheless a reason to attach some importance to these studies is the fact that they used in-depth and partly observational methods to study school and classroom processes and used appropriate state of the art methodology (econometric and multi-level modeling, use of achievement tests that allowed for vertical equating in the case of the study by Glewwe et al.).
The results are somewhat disconcerting with respect to the assumption that school organizational and particularly classroom level instructional variables will account for a sizeable part of the variance in achievement outcomes, as could be expected (for the class level conditions) on the basis of research in industrialized countries. Apparently an earlier review, by Anderson, Ryan and Shapiro (1989), reached a similar conclusion stating that

"variations in teaching practice in developing countries are only rarely found to be associated with variations in student learning" (Anderson et al., 1989, cited by Glewwe et al., 1995).
The 1994 review by Fuller and Clarke, however, indicate significant effects for a large proportion of the limited occasions when classroom pedagogy and school management variables were included (see table 10).

Two lines of reasoning could be considered to explain a relatively low impact of organizational and instructional variables.

- The first is the reference to cultural contingencies put forward by Fuller and Clarke, already discussed in a previous section.
- The second calls upon bearing in mind the comparative nature of the school effectiveness studies that were reviewed. It could well be that, in some developing countries, the range of variation in teaching practices is quite limited.

The implications for future research of the first would be the exploration and measurement of the most relevant cultural contingencies, preferably in internationally comparative studies. With respect to the second supposition, a more focused study of the between school variances on relevant process indicators should be considered. Such an approach is in line with the identification of process indicators as carried out by Heneveld & Craig (1994).

More recent contributions to the study of instructional effectiveness

In this section the results of some more recent contributions and reviews will be presented, on the basis of work by Anderson, 1991, 2004, Brophy, 2001, Baumert et al., 2000, NCS, 2002, Muijs and Reynolds, 2001, OECD, 2003. In these reviews, a strong corroboration of the main characteristics of effective instruction as laid out in the previous sections can be discerned. In addition to this consolidation in the knowledge base there are a few additional newer trends. These are the following:

- a reconsideration of personal characteristics of effective teachers
- more attention to the teaching of higher order skills, self-regulated learning and "constructivist" approaches
- a strong re-statement of the fact that teaching is about facilitating learning, by considering learning activities and student engagement

In the United States the issue of effective teacher characteristics is receiving much attention in the debate about standards for teaching competency (Darling Hammond, 2000). Empirical studies indicate that subject matter mastery and verbal skills are important assets of teacher effectiveness. In the United Kingdom, Hay McBerr (2000, cited by Anderson, 2004) identified twelve characteristics, in the sense of relatively stable traits, associated with effective teachers. These are stated in the table below, cited from Anderson, 2004, p. 15).

Table 4: Summary of Characteristics Associated with More Effective Teachers

Cluster	Characteristic	Description
Professionalism	Commitment	Commitment to do everything possible for each student and enable all students to be successful
	Confidence	Belief in ones ability to be effective and to take on challenges
	Trustworthiness	Being consistent and fair; keeping ones word
	Respect	Belief that all persons matter and deserve respect
Thinking/Reasoning	Analytical thinking	Ability to think logically, break things down, and recognize cause and effect
	Conceptual thinking	Ability to see patterns and connections, even when a great deal of detail is present
Expectations	Drive for improvement	Relentless energy for setting and meeting challenging targets, for students and the school
	Information Seeking	Drive to find out more and get to the heart of things; intellectual curiosity
	Initiative	Drive to act now to anticipate and preempt events
Leadership	Flexibility	Ability and willingness to adapt to the needs of a situation and change tactics
	Accountability	Drive and ability to set clear expectations and parameters and hold others accountable for performance
	Passion for Learning	Drive and ability to support students in their learning, and to help them become confident and independent learners

Note. Adapted from Hay McBer, (2000) by Anderson, 2004.

An interesting feature in this list is the "drive for improvement". This trait is similar to the "relentlessness" that is emphasized in Slavin's "Success for All" program, 1999) and what Anderson and Pellicer (1998) have called "zero tolerance to failure". The dimension of confidence is associated with the "high expectations" factor in the school and classroom climate, as one of the frequently identified factors of effective schooling. It seems very likely that these motivationally oriented personal characteristics of teachers have an impact on climate and culture of the school and the classroom.

Anderson's review can be seen as a re-statement and summary of results that have been presented in the earlier sections of this chapter. A specific asset of his monograph is the fact that it contains many examples of items and instruments that can be used to measure the central concepts of teacher effectiveness.
The basic framework from Anderson's work is given in the figure below.

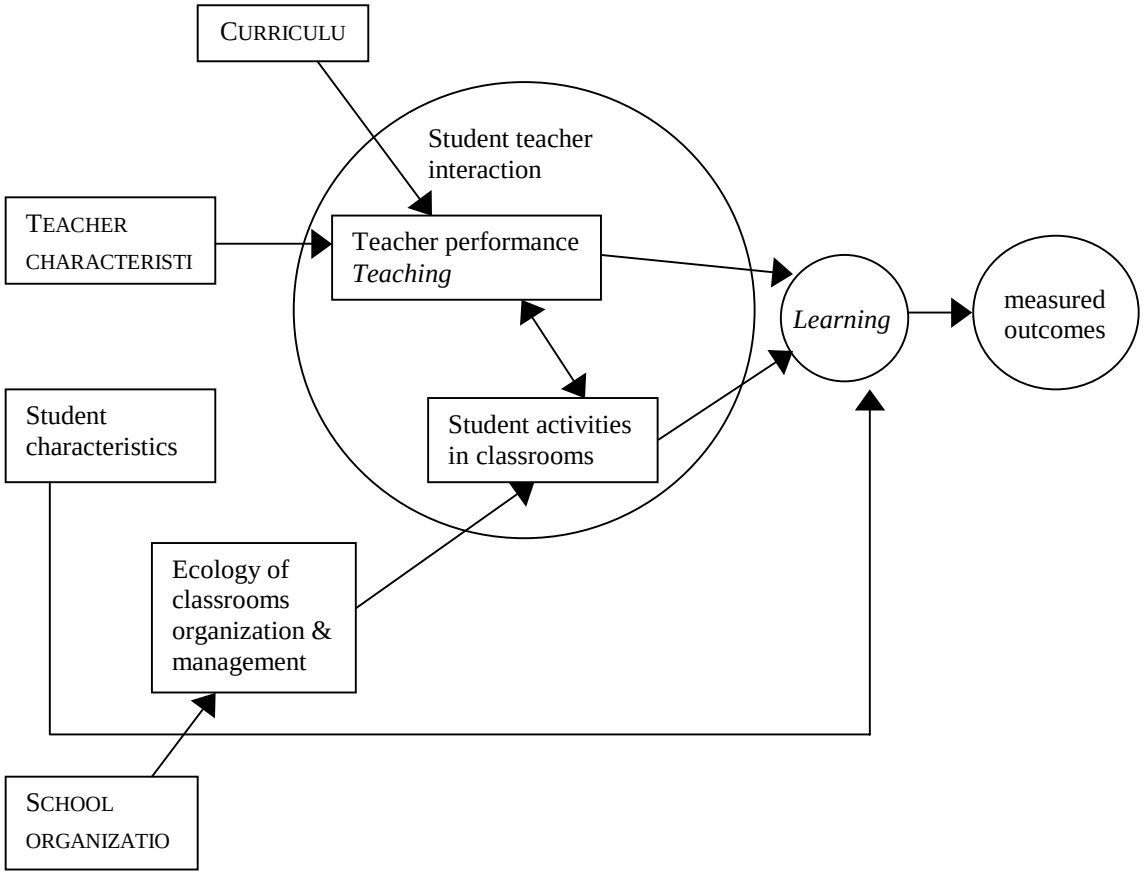


Figure 1: *Conceptual model on effective teaching, adapted from Anderson, 1991.*

The central category of *teaching* is subdivided in four major areas: the enacted curriculum, (the substantive curricular choices that a teacher makes), classroom environment and climate (in the sense of the physical environment and the psychological environment), actual teaching activities (the way lessons are structured and aspects of communication between teacher and students) and stimulating involvement.
In Table 3 below the key variables within these categories are included.

Table 3: *Conditions of effective teaching according to Anderson, 1999.*

Teaching characteristics (Anderson)	Teaching
enacted curriculum • opportunity to learn • academic work: - appropriate selection - regular (home)work - students held accountable - adequate supervision	*structuring of lessons (sequence)* • orientation • clear purposes • monitoring • independent practice • corrective feedback *communication teachers/students* • clear explanations • showing & telling • appropriate guiding • providing feedback
Classroom environment and climate	Stimulating involvement
physical environment • classroom arrangement • equipment • seating patterns • class size *climate (psychological environment)* • mutual respect • task orientation • structure	• reinforce paying attention • develop learning strategies instruments • success standards • create “holding power” in learning • keep students actively involved • circulate during seatwork • communicate interest

Clearly, factors like opportunity to learn, time on task, a structured approach to teaching, feedback and reinforcement that stood out from the earlier studies and reviews are also central in Anderson’s study.

Brophy’s work has been referred to in the earlier sections of this paper. The review he did with Good, (in 1979) is a classic in the literature on instructional effectiveness. In a more recent review, published in Wang and Walberg, 2001, he distinguishes 12 principles of effective teaching. These are the following:

1) Supportive classroom climate: students learn best within cohesive and caring learning communities. The role of the teacher as models and socializers is emphasised
2) Opportunity to learn: students learn more when most of the available time is allocated to curriculum related activities and the classroom management system emphasises maintaining students’ engagement in those activities.
3) Curricular alignment: All components of the curriculum are aligned to create a cohesive program for accomplishing instructional purposes and goals.

4) Establishing learning orientations: teachers can prepare students for learning by providing an initial structure to clarify intended outcomes and cue desired learning strategies. (e.g., providing advance organisers and cueing the kind of responses that are expected).
5) Coherent content: to facilitate meaningful learning and retention, content is explained clearly and developed with an emphasis on its structure and connections. " When making presentations, providing explanations, or giving demonstrations, effective teachers project enthusiasm for the content and organize and sequence it so as to maximise its clarity and "learner friendliness".
6) Thoughtful discourse: questions are planned to engage students in sustained discourse structured around powerful ideas.
7) Practice and application activities: students need sufficient opportunities to practice and apply what they are learning, and to receive improvement-oriented feedback.
8) Scaffolding students' task engagement: the teacher provides whatever assistance students' need to enable them to engage in learning activities productively. Structuring and support can be lessened as the students' expertise develops.
9) Strategy teaching: the teacher models and instructs students in learning and self-regulation strategies. Meta-cognitive awareness and self-regulation are sought in context like problem solving and general learning and study skills. An example of teacher modelling is, for example, when a teacher thinks out loud while modelling use of the strategy. Students are stimulated to monitor and reflect on their learning.
10) Co-operative learning: students often benefit from working in pairs or small groups to construct understandings or help one another master skills.
11) Goal-oriented assessment: The teacher uses a variety of formal and informal assessment methods to monitor progress toward learning goals. Comprehensive assessment also examines students' reasoning and problem-solving processes.
12) Achievement expectations: the teacher establishes and follows through on appropriate expectations for learning outcomes.

It is interestingly to note that quite a few of Brophy's principles are variations on the theme of structured teaching (advance organisers, stating clear goals, scaffolding, frequent monitoring, feedback). The next interesting point is the incorporation of some ideas from constructivism: attention for modelling, self-regulated learning as well as meta-cognitive processes.

Baumert et al. 2000, interpret instruction as an opportunity structure for insightful learning. "This means that instructional materials, task selection, and instructional processes are analyzed from the perspective of whether they foster or obstruct active individual knowledge acquisition. …. Dimensions of this opportunity structure include the safeguarding of the social action framework by means of appropriate classroom management; pacing and range of learning opportunities (quantity of instruction); general instructional quality, in particular the didactical quality of the structure and realization of the instruction; and the quality of teacher-student and student-student relations."
These authors go on to say that:

“With respect to general properties of classroom management and the quantity and quality of instruction, robust findings from previous research give a good indication of which aspects of mathematics instruction need to be assessed..” (in the context of the OECD PISA-study)

“- Important properties of *classroom management* include: clearly defined rules and procedures, prevention of disturbances, effective responses to critical events, and routinization of basic social acts in the classroom.
- Important aspects of *learning opportunities* and pacing (and hence the quantity of instruction) include: learning opportunities with respect to the test items, appropriate ratio of material covered to lesson time (pacing), faithfulness to objectives and relevance of the instructional materials, pressure to perform and interaction tempo.
- The basic properties of *instructional quality* include: level of difficulty, clarity and structure in the presentation of material, adaptivity and individualization of instruction, remediality, participation in instructional activities, monitoring of student activities, and general constructivist properties of insightful learning “.

The following dimensions of the quality of teacher-student relations are considered: the teacher's ability to motivate students, social orientation, and diagnostic competence in the social domain, as well as the students' general satisfaction with their subject teacher.
About the quality of student-student relations in learner groups they propose the following dimensions: cohesion and formation of cliques, competitiveness, mutual assistance, aggression, and violation of norms. They also underline the significance of subjective norms with respect to the academic or non-academic orientation of the student body.
Again, in this contribution there is a clear integrative approach with respect to the “traditional” aspects of structured teaching and constructivist ideas on learning and instruction. (Baumert et al., ibid)

A final recent review on instructional effectiveness is the one carried out by the US National Center for Education Statistics (NCS), NCS, 2002. The focus of this report is a critical review of survey based methods to capture instructional processes. The conclusion is that survey based methods have failed to account sufficiently for the engagement of students with content although they have been successful in assessing the extent to which teachers deliver appropriate content. The report goes on to state bluntly that “measuring student engagement requires observational methods” with assessing teachers’ substantive and pedagogical knowledge as a second methodological option.

There is one other dimension in which more recent contributions return to an aspect that was also present in the very early publications on teaching effectiveness, as Gagne’s conception about “the conditions of learning” (Gagne, 1972) and the Caroll model, (Caroll, 1963). This is attention for student engagement and learning strategies as the ultimate “mediator” between teaching activities and student outcomes. In the OECD study on “student approaches to learning” a range of variables related to engagement is discerned, variables like “self-efficacy”, “instrumental motivation” and subject matter interest (OECD, 2003). As learning

strategies a distinction is made between memorization strategies, elaboration strategies and control strategies. Control strategies have a significant association with reading performance. Control strategies refer to students' ensuring that their learning goals are reached. "These strategies involve checking what one has learned and working out what one still has to learn, allowing learners to adapt their learning to the task at hand", (OECD, 2003, p.13). In a way these control strategies are the pendant of the main features of "structured teaching" and direct instruction, where it is the teacher who actively orders and controls the teaching and learning situation. When putting these two orientations, structured teaching on the one hand, and students effectively employing control strategies next to one another the following types of associations can be discerned:

- structured teaching happens as a *substitute* for student control strategies
- structured teaching happens as an *additional support* for student control strategies
- structured teaching happens as a *model and example* to enhance student control strategies
- structured teaching happens as a *suppressor* of student control, because students are not given sufficient leeway to develop and manifest this behavior themselves.

Weaker students in primary and secondary education are more likely to benefit from the first two alternatives, whereas the last two alternative combinations are more probable when dealing with better students in secondary education (where obviously the third alternative is a positive and the fourth a negative example).

Anderson, 2004, has stated a set of recommendations to support learning and stimulate student engagements, these are summarized in Table 5, below.

Table 5: *Recommendations to stimulate student engagement and support learning, from Anderson, 2004*

- During presentations, teachers should use techniques that keep students attentive and involved.
- Teachers should select assignments based not only on the relevance of the work to the lesson and unit objectives, but the appropriateness of the work of the students.
- When giving assignments, teachers should set standards for success and let students know they are all expected to meet them.
- Students should be held accountable for both the completion and the quality of their work. In turn, teachers should ensure that students have learned reasonably well what they are expected to learn before they are allowed to work on their own.
- Teachers should use the knowledge they gain from their students' work not only to evaluate students, but also to provide students with additional instruction targeted toward their errors and misunderstandings.
- During seatwork (and group work), teachers should circulate among students and monitor their work.
- Teachers should reinforce students for paying attention and expending the effort needed to learn and learn well.
- Teachers should regularly monitor student involvement in learning.

The results of these more recent reviews, in the sense of the most important instructional conditions that were referred to, are summarised in the table 6 below. The table includes the main observation categories of a classroom observation schedule that is currently being used in an internationally comparative study of SICI, an international organisation of educational Inspectorates in Europe.

Table 6: *Summary of recent reviews and the observation categories of the Dutch Inspectorate*

Teaching (Anderson)

enacted curriculum
classroom physical environment
classroom climate
classroom organisation & management
actual teaching
- pre-conditions (lesson planning)
- communication with students
- stimulating involvement

Brophy

opportunity to learn
curricular alignment
supportive classroom climate
achievement expectations
cooperative learning
goal-oriented assessment
coherent content; clear explanations
thoughtful discourse
establishing learning orientations
sufficient opportunities for practice and application
scaffolding student's task engagement
modeling learning and self-regulation strategies

Dutch inspectorate

- learning time
- support in climate
- challenge in climate
- structure in teaching
- activating students
- teaching learning strategies
- attainment/teacher focus on attention
- classroom organization

Baumert et al.

quantity and quality of instruction
teacher student relations
student student relations

Conclusion: how can effective teaching be stimulated by educational policy?

When summarising the results of PART II of this review of instructional effectiveness one could first of all re-state the conclusion that learning outcomes depend on student learning strategies and student motivation to learn (engagement). In turn several teaching conditions appear to be important to steer student learning to the attainment of the desired outcomes, these could be summarised in terms of three major dimensions: relevance, time, and structure.

The first dimension is about keeping subject matter selection on target and about alignment in the curriculum domain. It has to do with gearing the subject matter and other contents that is to be taught to standards and curriculum objectives. It includes assuring that the implemented curriculum matches the intended curriculum, taking care of curriculum alignment between grades and classes, and assuring that the contents of teaching and learning assignments match the contents of tests and other assessment instruments. This latter aspect is usually described as "opportunity to learn".

The second dimension starts from the time that is to be spent on major curriculum areas and subjects according to official timetables. It becomes further specified at school level in the sense of specific school priorities and the "net teaching time" that occurs in the classroom, and which could be defined as the official teaching time minus time "lost" to other activities. Ultimately time accumulates in "time on task", that is the time that students are actively involved in

learning activities.

The third dimension is the teaching "technology" (in the procedural sense, and not necessarily in the sense of applying media and information technology). Structured teaching, as in forms like "direct teaching" and "mastery learning" has received a lot of support in instructional effectiveness studies, particularly for disadvantaged learners in primary schools, but also at higher levels of schooling and with respect to higher level cognitive skills. On the basis of insights inspired by constructivism this dimension should perhaps better be seen as a continuum running from providing a lot of structure and "scaffolding" to the learning process, to giving independence to learners. Effective teaching could then be seen as applying the right level of structure given relevant characteristics of learners, learning tasks and educational objectives. The structure dimensions include frequent monitoring of students' progress and providing feedback and reinforcement related to the assessment outcomes. In this sense it is not only about providing cognitive support, but also about fostering student engagement. Adapting difficulty levels to the specific needs of the students can also be seen as a specific aspect of structuring teaching.

Other aspects of the teaching and learning situation like the classroom climate and classroom organisation have an impact on student learning that is sometimes also quite direct, as with respect to aspects of the school climate, like mutual trust, safety, discipline and achievement orientation. On the other hand these factors also appear to have a less direct impact, for example as when an efficient classroom organisation safeguards teaching time. An important aspect of the overall climate and environment is the composition of the student body in a school or classroom

Other factors that were dealt with in the review, like the characteristics of teachers, are seen as having a more indirect effect on students' learning activities and learning outcomes. Teacher characteristics could be seen as prerequisites for delivering relevant teaching, using time efficiently and providing the appropriate level of structure. Conditions of effective schooling, as treated in the first part of this chapter, like the school organisation and educational leadership, in their turn are seen as supportive conditions of effective teaching.

The results of this review, in the sense of a listing of the most important effectiveness enhancing teaching conditions are presented in Table 7 below.

<u>Relevance</u>	Opportunity to learn Curriculum alignment
<u>Time</u>	Learning time
<u>Structure</u>	Structured teaching Stimulating engagement Monitoring and questioning

	Feedback and reinforcement Modelling learning/self-regulation
Classroom environment	Task-oriented climate Mutual respect Orderliness, safety
Teacher characteristics	Subject matter mastery Verbal intelligence Teaching repertoire Achievement orientation

The final question is how (national) educational policy could stimulate teacher and teaching effectiveness. This is not a simple matter. In Western cultures teachers have traditionally enjoyed at lot of professional autonomy. In the organisational model of the school as a "professional bureaucracy" schools are described as "flat" organisations, with little need for management. In fact, according to this model, teacher training is the most important mechanism to control and co-ordinate teaching. Clearly initial training and in-service training remain important means to improve teacher competencies. In more recent ideas about schools as professional learning cultures and schools as learning organisation the idea of teachers as more permanent learners is propagated. The next major form of national control over teaching is the curriculum, particularly when it is standard based. In situations where the national curriculum is less specified, textbooks can take over this role, but then the role of national policy planners could only exist when these books are produced by the central administration, not, of course when this is dependent on the market. In other areas of enhancing effectiveness there is a clear role for the central administration to create favourable conditions, particularly to provide well-equipped school buildings, material resources and textbooks.

Currently there appear two major perspectives on enhancing teacher and teaching effectiveness through national policy measures in the curriculum domain.

The first is the one that Muijs and Reynolds (2001) call the "ownership paradigm". According to this paradigm "it was believed that teachers would be more likely to be effective and to develop as professionals if they were involved in actually creating the methods that their schools and classrooms would then reflect" (ibid p.217). This "bottom up" approach has also been quite popular for the last 15 years among people working in the field of school improvement in the USA the UK and the Netherlands. It appeals to the philosophy of decentralisation and stimulating school autonomy, and in political systems that are against "state pedagogy" and for curricula that only provide general frames of aims and methods. As Muijs and Reynolds say, the "ownership" paradigm has led to a lack of consistency in teaching approach between teachers, and a lot of "re-inventing the wheel". Contrary to the currently fashionable belief in the before mentioned countries teachers, if given the choice, have been seen to prefer "off the shelf" explicit teaching methods to promote educational effectiveness. The success of the "Success for

All" program in the USA also seems to depend partly on the almost "teacher proof" teaching guidelines (Slavin, 1999).

The second approach can be characterised as a pro-active centralised curriculum strategy. This strategy would include a standard based national curriculum in the main subject matter areas. Moreover the strategy would ensure and monitor alignment aspects, in the sense that curricula are operationalised into actual teaching programs, thus enlarging the chances of a good match between the official and the national curriculum. Explicit guidelines for teachers on "how to teach" the contents are part of these programs. The third pillar of this strategy consists of a set of tests, assessment or examinations to monitor the actual teaching effects. From the point of view of using the knowledge base on educational effectiveness this approach has the advantage that insights on which there is considerable consensus could be implement and made available to a large population of teachers. One important side issue would be the ability to monitor "opportunity to learn", another would be to gear teacher training and in-service training closely to the aims and methods set out in the national curriculum. Opportunity could be monitored because there would be sufficient clarity on the intended curriculum on the one hand, and the "realised curriculum" (i.e. the assessment instruments) on the other to assess the degree to which these are matched by what is actually taught at school. It would be a caricature to see such a curriculum as totally prescriptive. Even when clear guidelines are being provided there is always need for the professional skills of teachers to adapt to the students in the classroom and the students at hand.

A variation of this curriculum centralisation strategy would be a more retroactive interpretation, in which the assessment program would actually steer teaching priorities. In order to preclude non-desirable effects as pure "training for the test" this would put strong demands on the quality of the assessment instruments. These could actually provided by means of item banking techniques and tests that conform to Item Response models.

In countries in which the level of teacher training is problematic centralist curriculum strategies should be considered a better option for fostering teacher and teaching effectiveness. [1]

References

Achilles, C.M. (1996). Students achieve more in smaller classes. *Educational Leadership*, 76-77.

Aitkin, M., & Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Association*, Series A (General), 149, Part 1, 1-43.

Aldridge, B.G. (1983). A mathematical model for mastery learning. *Journal of Research in Science Teaching, 20*, 1-17.

[1] For an extended discussion on the issue of initiating educational change and school improvement see Annex 2 .

Alexander, K.L., & Eckland B.K. (1980). The "explorations in quality of opportunity" sample of 1955 high school sophomores. In A.C. Kerkhoff (Ed.*), Research in sociology of education and socialization*. Vol. I: Longitudinal perspectives on educational attainment. Greenwich: JAI Press Inc.

Anderson, C.S. (1982). The search for school climate: A review of the research. *Review of Educational Research, 52*(3), 368-420.

Anderson, L.W. (1991). *Increasing Teacher Effectiveness*. Paris: UNESCO, IIEP.

Anderson, L.W., (2004) *Increasing Teacher Effectiveness.* (Second Edition) Paris: UNESCO, IIEP

Anderson, L.W., & Pellicer, L.O., (1998) Towards an Understanding of Unusually Successful Programs for Economically Disadvantaged Students. *Juornal of Education for Students Placed at Risk, 3, 237- 263*

Anderson, L.W., Ryan, D.W., & Shapiro, B.J. (1989). *The IEA Classroom Environment Study*. Oxford: Pergamon Press.

Averch, H.A., Carroll, S.J., Donaldson, T.S., Kiesling, H.J., & J. Pincus (1974*). How effective is schooling? A critical review of research*. Englewood Cliffs, NJ: Educational Technology Publications.

Bangert, R.L., Kulik, J.A., & Kulik, C.C. (1983). Individualized systems of instruction in secondary schools. *Review of Educational Research*, 53, 143-158.

Baumert, J., Blum, W., & Neubrand, M (2000). Surveying the Instructional Conditions and Domain-Specific Individual Prerequisites for the Development of Mathematical Competences. (Draft paper)

Bereiter, C., & Kurland, M. (1982). A constructive look at follow through results. *Interchange*, 12, 1-22.

Berliner, D.C. (1985). Effective classroom teaching: the necessary but not sufficient condition for developing exemplary schools. In G.R. Austin & H. Garber (Eds.), *Research on Exemplary Schools*, pp. 127-154. Orlando, Fl.:Academic Press.

Block, J.H., & Burns, R.B. (1970). Mastery learning. *Review of Research in Education, 4*, 3-49.

Bloom, B.S. (1968). *Learning for Mastery*. Washington, D.C: ERIC.

Bloom, B. (1976). *Human characteristics and school learning*. New York: McGraw Hill.

Boekaerts, M., & Simons, P.R.J. (1993). *Leren en Instructie. Psychologie van de Leerling en het Leerproces*. Assen: Dekker & Van de Vegt.

Bosker, R.J., & Scheerens, J. (1999). Openbare prestatiegegevens van scholen; nuttigheid en validiteit. *Pedagogische Studiën, 76*(1), 61-73.

Bosker, R.J., & Witziers, B. (1996). The magnitude of school effects. Or: Does it really matter which school a student attend. Paper presented at AERA Annual meeting, New York.

Brandsma, H.P. (1993). *Basisschoolkenmerken en de kwaliteit van het onderwijs.* Groningen: RION.

Brookover, W.B., Beady, C., Flood, P. et al. (1979). *School social systems and student achievement - schools can make a difference*. New York: Praeger Publishers.

Brophy, J. (1996). *Classroom management as socializing students into clearly articulated roles*. Paper presented at AERA Annual Meeting, New York.

Brophy, J. (2001) Generic Aspects of Effective Teaching. In Wang, M.C. & Walberg, H.J. *Tomorrow's Teachers*. McCutchan Publishing Company.

Brophy, J., & Good, T.L. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.), *Third Handbook of Research on Teaching* [pp. 328-375]. New York: MacMillan.

Bruner, J.S. (1966). *Towards a Theory of Instruction*. Cambridge, Mass.: Belknap Press

Card, D. & Krueger, A.B. (1992). Does school quality matter? Returns to education and the characteristics of public schools in the United States. *Journal of Political Economy*, 100, 1-40.
University.

Carroll, J.B. (1963). A model of school learning. *Teachers College Record, 64*, 722-733.

Carroll, J.B. (1989). The Carroll Model, a 25-year retrospective and prospective view. *Educational Researcher, 18*, 26-31.

Cohen, M. (1982). Effective schools: Accumulating research findings. *American Education*, January-February, 13-16.

Cohen, D.K. (1988). Teaching practice ... Plus ça change ... In Ph. Jackson (ed.), *Contributing to Educational Change: Perspectives on Research and Practice*. Berkeley, CA.: McCutchan.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.

Collins, A., Brown, J.S. & Newman, S.E. (1989). Cognitive apprenticeship: teaching the crafts of reading, writing and mathematics. In L.B. Resnick (Ed.) *Knowing, learning and instruction* (pp.453-495) Hillsdale: Lawrence Erlbaum.

Collins, A., & Stevens, A. (1982). Goals and strategies of inquiry teachers. *Advances in Instructional Psychology*, Vol. II. Ed.: R. Glaser. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Corte, E. De, & Lowyck, J. (1983). Heroriëntatie in het onderzoek van het onderwijzen. [Research on teaching reconsidered]. *Tijdschrift voor Onderwijsresearch, 8*(6), 242-261.

Cotton, K. (1995). *Effective schooling practices: A research synthesis*. 1995 Update. School Improvement Research Series. Northwest Regional Educational Laboratory.

Creemers, B.P.M. (1994). *The Effective Classroom*. London: Cassell.

Creemers, B.P.M. (1996). *Common elements in theories about effective instruction*. Groningen: unpublished manuscript.

Cunningham, D.J. (1991). In defense of extremism. *Educational Technology, 31*(9), 26-27.

Darling-Hammond, L. (2000). Teacher Quality and Student Achievement: A Review of State Policy Evidence. *Education Policy Analysis Archives*, 8(1). Available: http://olam.ed.asu.edu/epaa/v8n1/.

Davies, J.K. (1972). Style and effectiveness in education and training: a model for organizing, teaching and learning. *Instructional Science, 2*.

Dougherty, K. (1981). After the fall: Research on school effects since the Coleman report. *Harvard Educational Review*, 51, 301-308.

Doyle, W. (1985). Effective secondary classroom practices. In: M.J. Kyle (Ed.), *Reaching for excellence. An effective schools sourcebook.* Washington: U.S. Government Printing Office.

Duffy, Th.M., & Jonassen, D.H. (1992). *Constructivism and the Technology of Instruction: A Conversation*. Hillsdale, NJ: Lawrence Erlbaum Ass.

ECD (2003)

Edmonds, R.R. (1979). Effective schools for the urban poor. *Educational Leadership, 37*(1), 15-27.

Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation, 28*(2), 465–498.

Fraser, B.J., Walberg, H.J., Welch, W.W., & Hattie, J.A. (1987).Syntheses of educational productivity research. Special Issue of *the International Journal of Educational Research, 11*(2).

Fuller, B. (1987). What factors raise achievement in the Third World? *Review of Educational Research*, 57, 255-292.

Fuller, B., & Clarke, P. (1994). Raising school effects while ignoring culture? Local conditions and the influence of classroom tools, rules and pedagogy. *Review of Educational Research*, 64, 119-157.

Fuller, B., Hua, H., Snijder, C.W. (1994). Focus on Gender and Academic Achievement. When Girls learn more than Boys: The Influence of Time in School and Pedagogy in Botswana. *Comparative Education Review*, 38, 347-376.

Gage, N. (1965). Desirable behaviors of teachers. *Urban Education, 1*, 85-95.

Gagne, R.M (1972) *Conditions of Learning.* New York: Holt, Rinehart and Winston.

Galton, M. (1999). Class Size and Pupil Achievement. *International Journal of Educational Research*, vol. 29.

Glasman, N.S., & Biniaminov, J. (1981). Input-output analysis of schools. *Review of Educational Research*, 51, 509-539.

Glewwe, P., Grash, M., Jacoby, H., Lockheed, M. (1995). An Eclectic Approach to Estimating the Determinants of Achievement in Jamaican Primary Education. *The Work Bank Economic Review*, 9, 231-258.

Good, Th.L., & Brophy, J. (1986). School effects. In: M.C. Wittrock (ed.), *Handbook of research on teaching* [p. 328-375]. New York: McMillan Inc.

Grisay, A. (1996). *Evolution des acquis cognitifs et socio-affectifs des eleves au cours des annees de college*. Liège: Université de Liège.

Hanushek, E.A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources*, 14, 351-388.

Hanushek, E.A. (1986). The economics of schooling: production and efficiency in public schools. *Journal of Economic Literature*, 24, 1141-1177.

Hanushek, E.A. (1989). Expenditures, efficiency, and equity in education: the federal Government's role. *The American Economic Review, 79*(2), p. 46.

Hanushek, E.A. (1995). Interpreting Recent Research on Schooling in Developing Countries. *The World Bank Research Observer*, 10(227-46).

Hanushek, E.A. (1997). Assessing the effects of school resources on student performance: an update. *Educational Evaluation and Policy Analysis, 19*, 141-164.

Hauser, R.M., Sewell, W.H., & Alwin, D.F. (1976). High School effects on achievement. In: W.H. Sewell, R.M. Hauser & D.L. Featherman (Eds.), *Schooling and achievement in American Society*. New York: Academic Press.

Haywood, H.C. (1982). Compensatory education. *Peabody Journal of Education*, 59, 272-301.

Hedges, L.V., Laine, R.D., & Greenwald, R. (1994). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher, 23*, 3, 5-14.

Heneveld, W., & Craig, H. (1996). *Schools Count. World Bank Project Designs and the Quality of Primary Education in Sub-Saharan Africa*. Washington D.C.: World Bank.

Hill, P.W., Rowe, K.J., & Holmes-Smith, P. (1995). *Factors Affecting Students' Educational Progress: Multilevel Modelling of Educational Effectiveness*. Paper presented at the 8th International Congress for School Effectiveness and Improvement, Leeuwarden, the Netherlands, January 3-6, 1995.

Hoeven - Van Doornum, A.A. van der (1990). Effecten van leerlingbeelden en streefniveaus op schoolloopbanen [The impact of teachers' expectations on educational careers]. Nijmegen: ITS/OOMO.

Jencks, C., Smith, M.S., Ackland, H., Bane, M.J., Cohen, D., Grintlis, H., Heynes, B., & Michelson, S. (1972). *Inequality. A reassessment of the effect of family and schooling in America.* New York: Basic Books.

Jencks, C. et al. (1979). *Who gets ahead? The determinants of economic success in America*. New York: Basic Books.

Jimenez, E. & Paquea, V. (1996). Do Local Contributions Affect the Efficiency of Public Schools? *Economics of Education Review*, 15, 377-386.

Johnston, K.L., & Aldridge, B.G. (1985). Examining a mathematical model of mastery learning in a classroom setting. *Journal of Research in Science Teaching*, *22*(6), 543-554.

Jonassen, D.H. (1992). Evaluating constructivist learning. In Th.M. Duffy & D.H. Jonassen (eds.), *Constructivism and the Technology of Instruction: A Conversation*, 138-148. Hillsdale, NJ: Lawrence Erlbaum Ass.

Karweit, N., & Slavin, R.E. (1982). Time on task: issues of timing, sampling and definition. *Journal of Educational Psychology*, *74*, 844-851.

Klerk, L.F.W. de (1985). ATI-onderzoek en differentiatie: een reactie. *Pedagogische Studiën*, *62*, 372-375.

Kulik, C.L.C., & Kulik, J.A. (1982). Effects of ability grouping on secondary school students: a meta-analysis of research findings. *American Educational Research Journal*, *19*, 415-428.

Kyle, M.J. (Ed.) (1985). Reaching for excellence. An effective schools sourcebook. Washington D.C.: U.S. Government Printing Office.

Laarhoven, P. van, & Vries, A.M. de (1987). Effecten van de interklassikale groeperingsvorm in het voortgezet onderwijs. Resultaten van een literatuurstudie. (Effects of heterogeneous grouping in secondary schools). In: J. Scheerens & W.G.R. Stoel (red.), Effectiviteit van onderwijsorganisaties. Lisse: Swets & Zeitlinger.

Levine, D.K., & Lezotte, L.W. (1990*). Unusually Effective Schools: A Review and Analysis of Research and Practice*. Madison, Wise: Nat. Centre for Effective Schools Research and Development.

Lockheed, M. & Hanushek, E. (1988). Improving educational efficiency in developing countries: What do we know? *Compare*, 18(1), 21-38.

Lockheed, M.E., & Longford, N.T. (1989). *A Multi-Level Model of School Effectiveness in a Developing Country*. Discussion paper no. 69. Washington D.C.: World Bank.

Lockheed, M. & Verspoor, A. (1991). *Improving Primary Education in Developing Countries*. London: Oxford University Press.

Lortie, D.C. (1973). Observations on teaching as work. In: R.M.W. Travers (Ed*.), Second Handbook of Research on Teaching*. Chicago: Rand McNally.

Medley, D., & Mitzel, H. (1963). Measuring classroom behavior by systematic observation. In: N.L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand McNally.

Merrill, M.D. (1991). Constructivism and instruction design. *Educational Technology*, *31*, 45-53.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988*). The junior school project; technical appendices*. London: ILEA, Research and Statistics Branc.

Mosteller, F., & D.P. Moynihan (Eds.). (1972*) On Equality of Educational Opportunity*. New York: Random House.

Muijs & Reynolds (2001)

Murmane, R.J. (1981). Interpreting the evidence on school effectiveness. *Teachers College Record*, 83, 19-35.

National Center for Education Statistics (NCS) (2002). *Fundamental problems in the measurement of instructional processes.* Washington D.C.: NCS.

Neufeld, E., Farrar, E., & Miles, M.B. (1983). A review of effective schools research: The message for secondary schools. National Commission on Excellence in Education.

Nyagura, L.M., & Riddell, A. (1993). *Primary School Achievement in English and Mathematics in Zimbabwe. A multilevel analysis.* The World Bank, Policy Research Working Paper.

OECD (2003).*Learners for Life. Student Approaches to Learning.* Results from PISA 2000, Paris: OECD

Picciotto, R. (1996). What is Education Worth? From Production Function to Institutional Capital. World Bank: Human Capital Development Working Paper, no. 75.

Postlethwaite, T.N., & Ross, K.N. (1992). *Effective Schools in Reading. Implications for Educational Planners*. The Hague: IEA.

Pritchett, L. & Filmer, D. (1997). What Educational Production Functions Really Show. A positive theory of education. The World Bank. Policy Research Working Paper of the Development Research Group Poverty and Human Resources.

Purkey, S.C., & Smith, M.S. (1983). Effective schools: a review. *The Elementary School Journal*, 83(4), 427-452.

Ralph, J.H., & Fennessey, J. (1983). Science or reform: some questions about the effective schools model. *Phi Delta Kappan*, 64(10), 689-695.

Reezigt, G.J. (1993). *Effecten van differentiatie op de basisschool* (Effects of grouping in primary education). Groningen: RION.

Resnick, L.B. (1987). *Education and Learning to Think*. Washington, DC: National Academic Press.

Reynolds, D., Hopkins, D. & Stoll, L. (1993). Linking school effectiveness knowledge and school improvement practice: towards a synergy. School Effectiveness and School Improvement, 4(1), 37-58.

Riddell, A. (1997). Assessing Designs for School Effectiveness Research and School Improvement in Developing Countries, *Comparative Education Review*, *41*(2).

Rosenshine, B.V. (1983). Teaching functions in instructional programs. *Elementary School Journal*, *3*, 335-351.

Rosenshine, B., & Furst, N. (1973). The use of direct observation to study teaching. In: R.M. Travers (Ed.), *Second Handbook of Research on Teaching*. Chicago: Rand McNally.

Rutter, M. (1983). School effects on pupil progress: research findings and policy implications. *Child Development*, 54(1), 1-29.

Sammons, P., Hillman, J., & Mortimore, P. (1995). *Key characteristics of effective schools: A review of school effectiveness research*. London: OFSTED.

Scheerens, J. (1989). Process indicators of school functioning. *School Effectiveness and School Improvement 1*(1), 61-80.

Scheerens, J. (1990). School effectiveness research and the development of process indicators of school functioning. *School Effectiveness and School Improvement*, 1(1), 61-80. Lisse: Swets & Zeitlinger.

Scheerens, J. (1992). *Effective Schooling, Research, Theory and Practice*. London: Cassell.

Scheerens, J. (1999). Recent developments in school effectiveness research in the Netherlands. In: T. Townsend, P. Clarke, & M. Ainscow (Eds.), *Third Millenium Schools; a world of difference in Effectiveness and Improvement* [pp.143-159]. Lisse: Swets & Zeitlinger.

Scheerens, J., & Bosker, R.J. (1997). *The Foundations of Educational Effectiveness*. Oxford: Elsevier Science Ltd.

Scheerens, J., & Creemers, B.P.M. (1989). Conceptualizing school effectiveness. In: B.P.M. Creemers & J. Scheerens (Eds.), Developments in school effectiveness research. Special issue of *the International Journal of Educational Research*, (13)7. Oxford: Pergamon Press, p. 691-706.

Scheerens, J., Vermeulen, C.J.A.J., & Pelgrum, W.J. (1989). Generalizability of school and instructional effectiveness indicators across nations. In: B.P.M. Creemers & J. Scheerens (eds.), *Development in school effectiveness research*. Special issue of the *International Journal of Educational Research*, 13(7), 789-800. Oxford: Pergamon Press.

Slavin, R.E. (1987). *Cooperative learning: student teams*, 2nd edition. Washington D.C.: National Education Association Professional Library.

Slavin, R.E. (1996). *Success for all*. Lisse: Swets & Zeitlinger.

Spiro, R.J., Feltowich, P.J, Jacobson, M.J., & Caulson, R.L. (1992). Cognitive flexibility, constructivism and hypertext: random access instruction for advanced knowledge acquisition in ill-structured domains. In Th.M. Duffy & D.H. Jonassen (eds.), *Constructivism and the Technology of Instruction: A Conversation*. Hillsdale, NJ: Lawrence Erlbaum Ass.

Stallings, J. (1985). Effective elementary classroom practices. In: M.J. Kyle (Ed.), *Reaching for excellence. An effective schools sourcebook.* Washington, DC: US Government Printing Office.

Stallings, J., & Mohlman, G. (1981*). School policy, leadership style, teacher change and student behavior in eight schools*. Final report to the National Institute of Education, Washington D.C.

Stebbins, L.B., St. Pierre, R.G., Proper, E.C., Anderson, R.R., & Cerva, T.R. (1977). Education as experimentation: a planned variation model, Vol. IV-A. An evaluation of Follow Through. Cambridge, Mass.: Abt Associates Inc.

Stringfield, S.C. & R. Slavin (1992). Raising societal demands, high reliabilty organizations, school effectiveness, "Succes for All", and a set of modest proposals. Enschede, Address to the ICO-symposium.

Sweeney, J. (1982). Research synthesis on effective school leadership. *Educational Leadership*, 39, 346-352.

Thorndike, R.L. (1973). Reading comprehension education in fifteen countries. Stockholm: Alonqvist & Wiksell.

Tobias, S. (1991). An eclectic examination of some issues in the constructivist-ISP controversy. *Educational Technology*, *31*(9), 41-43.

Vedder, P.H. (1985). *Cooperative learning. A study on processes and effects of cooperation between primary school children*. The Hague: SVO.

Verstegen, D.A. & King, R.A. (1998). The Relationship Between School Spending and Student Achievement: A Review and Analysis of 35 Years of Production Function Research. *Journal of Education Finance*, 24, 243-262.

Walberg, H.J. (1984). Improving the productivity of American schools. *Educational Leadership, 41*, 19-27.

Wang, M.C., Haertel, G.D., & Walberg, H.J. (1993). Toward a knowledge base for school learning. *Review of Educational Research, 63*(3), 249-294.

Wang, M.C., & Walberg, H.J. (2001). *Tomorrow's Teachers*. McCutchan Publishing Company.

Weeda, W.C. (1986). Effectiviteitsonderzoek van scholen. In: J.C. van der Wolf & J.J. Hox (red.), *Kwaliteit van het onderwijs in het geding.* (About education quality). Publicaties van het Amsterdams Pedogogische Centrum, nr. 2. Lisse: Swets & Zeitlinger.

Werf, M. P.C. van der, Van der, Lubbers M.J. & Kuyper, H. (1999). De opbrengsten van het voortgezet onderwijs voor de leerlingen uit de VOCL-cohorten 89-1 en 93-1 [The outputs of secondary education 1989 and 1993]. Groningen: RION.

ANNEX: THE OBSERVATION INSTRUMENT OF THE DUTCH INSPECTORATE

LESSON OBSERVATION FORM DUTCH INSPECTORATE

School registration number:

GROUP : lesson/task : DATE:
SCHOOL YEAR : textbook :
TEACHER : part :
TEACHING ASSISTANT :

TLP 2	**Learning time**	1	2	3	4	**n.a**	**NOTES:**
2.4	realisation of instruction	0	0	0	0		
TLP 3	**School climate**						
A	**support**						
3.1	safe, functional learning environm.	0	0	0	0		
3.2	respect for pupils	0	0	0	0		
3.3	promote mutual respect	0	0	0	0		
3.4	support self-confidence	0	0	0	0		
3.6	implement rules	0	0	0	0		
B	**challenge**						
3.7	challenging learning environment	0	0	0	0		
3.8	stimulate independence/ own responsibility	0	0	0	0	0	
TLP 4	**Teaching strategy**						
A	**clear and structured**						
4.1	clear structure	0	0	0	0		
4.2	clear explanation	0	0	0	0	0	
4.3	check/feedback	0	0	0	0		
B	**activating**						
4.4	actively involve pupils	0	0	0	0	0	
4.5	activating teaching methods	0	0	0	0		
4.7	learning by cooperation	0	0	0	0	0	
4.8	pupil involvement	0	0	0	0		
C	**learning to use learning stra**						
4.9	concrete, recognisable situations/ application oriented tasks	0	0	0	0		
4.10	learning strategies	0	0	0	0	0	
4.11	interaction aimed at strategies	0	0	0	0	0	
4.12	encourage pupils to check their actions	0	0	0	0		
D	**attuning**						
4.13	to the class as a whole	0	0	0	0		
	to individual differences:						
4.14	instruction	0	0	0	0	0	
4.15	processing	0	0	0	0	0	
4.16	use of language	0	0	0	0	0	
E	**class organisation**						
4.17	efficiency	0	0	0	0		

ANNEX to REVIEW OF SCHOOL AND INSTRUCTIONAL EFFECTIVENESS RESEARCH CONTRIBUTION TO CHAPTER 3 OF THE 2004 EFA GLOBAL MONITORING REPORT

Jaap Scheerens, March, 2004

THE USE OF INTERNATIONAL COMPARATIVE ASSESSMENT STUDIES TO ANSWER QUESTIONS ABOUT EDUCATIONAL PRODUCTIVITY AND EFFECTIVENESS

Introduction

In this annex four different types of uses of international comparative assessment studies will be briefly described and illustrated:

- comparison of country mean scores on a particular achievement test
- analyzing between schools, between classes and between student variation
- separating the effects of "given" conditions and malleable factors
- answering questions about the effectiveness of specific school, context and classroom characteristics

In the final section some conclusions will be drawn about the possibilities and limitations of these international studies to answer questions about school effectiveness.

The comparison of country average scores as a measure of productivity

The most frequent use that is made of the results of internationally comparative assessment studies, as those carried out by the IEA and the OECD, is to compare the country mean scores on a particular achievement test. When standard errors are presented with these averages, differences between countries that are over 2 times the standard error indicate statistical significance.

The table below (TableA.1), representing the results of the OECD PISA study in reading literacy, carried out in 2000, is presented as an illustration. When countries differ about 8 points from one another the differences are statistically significant.
The PISA results illustrate the substantial difference between the higher and lower scoring countries.

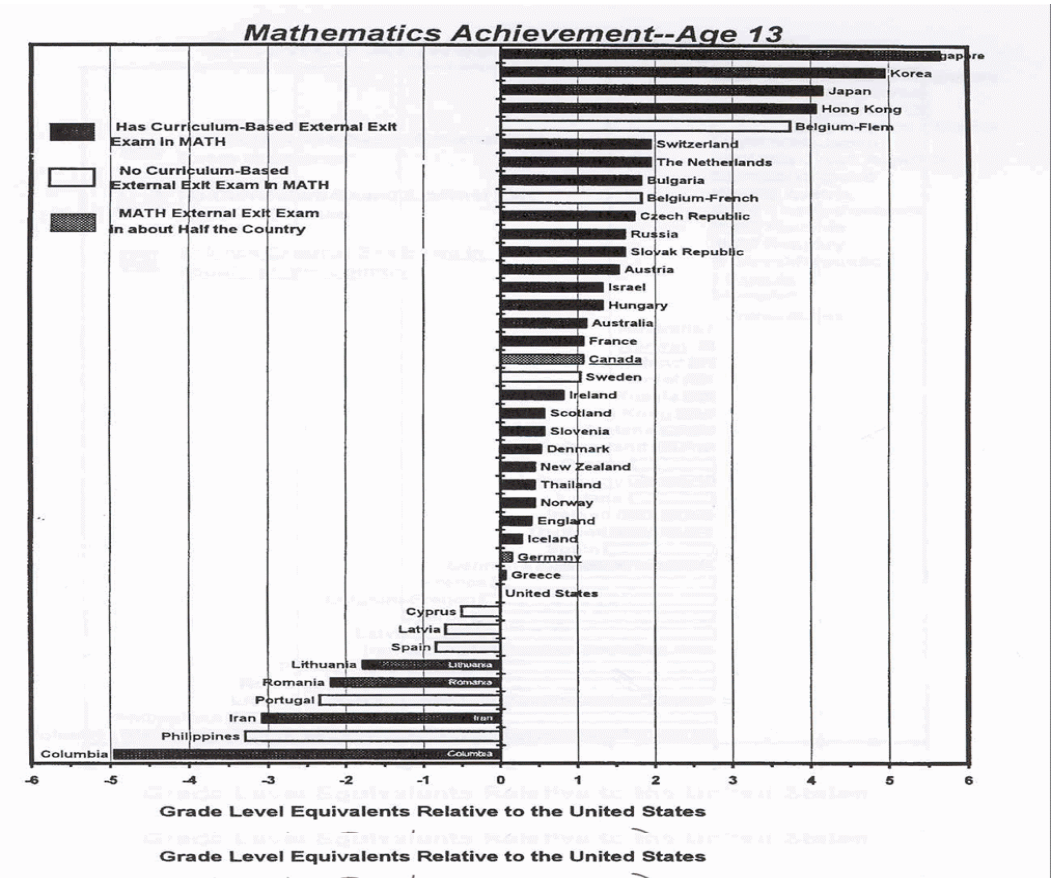
Data as presented in Table A.1 could be used as targets or benchmarks. Countries, for example might take the international average as a target for a future comparison.

Table A.1 Country averages on reading literacy. Source: PISA 2000 and PISA plus.

	Reading literacy score**
Finland	546
Canada	534
Netherlands	532
New Zealand	529
Australia	528
Ireland	527
Korea	525
Hong Kong	525
United Kingdom	523
Sweden	516
Austria	507
Belgium	507
Iceland	507
France	505
Norway	505
United States	504
Denmark	497
Switzerland	494
Spain	493
Czech Republic	492
Italy	487
Germany	484
Hungary	480
Poland	479
Greece	474
Portugal	470
Russian Federation	462
Latvia	458
Israel	452
Luxembourg	441
Thailand	431
Bulgaria	430
Mexico	422
Argentina	418
Chili	409
Brazil	396
Macedonia	372
Indonesia	371
Albania	349
Peru	322
Average across countries	473

Countries are ranked according to their average reading literacy score.

Table A.2: *Grade level equivalents relative to the United States (source: John H. Bishop, 1997)*



Grade Level Equivalents Relative to the United States

A perhaps more insightful way to present comparisons in achievement is to express differences in mean scores in terms of grade equivalents. A grade equivalent difference is defined as the difference in mean score between students at the beginning and the end of a particular grade level, in a particular country. In the table above, cited from Bishop, grade equivalents were defined on the basis of data from the USA. His table, by the way, contains also information on countries having a standard based examination system or not. The data are from TIMSS, 1995.

Analyzing between school, between classes, and total between student variation

International comparative studies allow for comparisons of the patterns of variation in student achievement scores. The between school, within country variation, is the percentage of total between student variation that is explained by the factor school, i.e. it expresses the difference it would make for the average student to be enrolled in one school as compared to the next. The between classes, within school variation expresses the difference it would make for an average student in that school to be in one parallel classroom (at the same grade level) as compared to the next. A large total between student variation in a country shows that there is much heterogeneity in student performance, which is likely to be interpreted as low equity.

A large between school variation expresses “implicit” of “explicit” segregation. The term implicit segregation could be used when, in a formal structural sense all schools are equal, as in the case of a comprehensive secondary education system. When there is still a large between school variation in a comprehensive system this could be caused by large school autonomy, or by selection policies of the schools or the parents that choose a school for their children. Explicit segregation appears when countries have a categorical school system, most common at secondary level. In categorical school systems students go to schools that cater to different ability levels.
Large between classes variation within schools, is indicative of within school tracking or streaming.

The table below, Table A.3 shows between school and between classes variation patterns in countries that took part in the Second Mathematics Study of the I.E.A. (source: Scheerens, Vermeulen & Pelgrum, 1989).

Table A.3: *Estimates of the Variance Explained by Schools and Classes*, cited from Scheerens et al. (1989)

Country	**Classroom variance component**	**School variance component**
15 Belgium (Flemish)		.50
16 Belgium (French)		.64
22 Canada (British Columbia)		.27
25 Canada (Ontario)	.18	.09
39 Finland	.45	.002
40 France	.17	.06
43 Hong Kong		.51
44 Hungary		.30
50 Israel	.22	.10
54 Japan		.08
59 Luxembourg	.29	.15
62 Netherlands		.67
63 New Zealand	.45	.01
72 Scotland	.34	.12
76 Sweden	.45	.00
79 Thailand		.39
81 USA	.46	.10

Note: Estimated of the variances expressed in terms of the intra-class correlation coefficient, for all countries, assuming schools are sampled at random within countries and classrooms are sampled at random within schools; the coefficients shown between brackets are the intra-class correlation coefficients are controlling for fathers' occupation

Since in 8 countries only one class per school was selected, classroom variance could not be separated from school variance in these cases. When looking at the results in table A.3 four groups of countries can be distinguished. First of all there are countries (Belgium Flemish, Belgium French, and The Netherlands) where there are vast differences in the mean achievement of students across schools: in this situation we have to do with vertically organized, strongly differentiated school systems. Secondly there is a group of countries with relatively small differences between schools but with large differences between classes within schools: the

USA, Sweden, New-Zealand, and Finland: this pattern indicates homogeneous grouping of pupils within a horizontally organized, integrated system of secondary schools. Next, there is a group of countries (Canada, France, Israel) where differences both between schools as well as between classes within schools are relatively small, probably because of (partially) mixed ability grouping within an integrated schooling system. Then of course there are countries that do not have a tracked, vertically organised system, but where de facto there are large quality differences between schools (most notably in Hong Kong and Thailand).

In a re-analyses of the PISA-2000 data set it appeared that different patterns can be discerned in countries having high versus low between school and total between student variation (Scheerens & Visscher, 2004). The following patterns appeared to occur:

High between school and high between students variation, e.g. **Germany**	Low between school and high between students variation, e.g. **New Zealand**
High between school and low between students variation, e.g. **Korea, the Netherlands**	Low between school and low between students variation, e.g. **Sweden**

The precise information is presented in Table A3 below.

Table A.3 (source Scheerens & Visscher, 2004)
The total variance and the proportion of variance at the school level in reading literacy scores of the students based on an empty model. The 95% confidence interval (CI) of the proportion of variance at school level* is also presented.

	Total variance	Proportion of variance at school level	Lower limit 95% CI	Upper limit 95% CI
OECD countries				
Australia	11407.42	0.21	0.16	0.27
Austria	9798.47	0.54	0.46	0.62
Belgium – French	12662.11	0.57	0.46	0.68
Belgium – Flemish	8601.80	0.52	0.43	0.61
Canada	9617.84	0.20	0.18	0.22
Czech Republic	8918.52	0.55	0.48	0.62
Denmark	9297.15	0.16	0.11	0.20
Finland	7640.28	0.07	0.04	0.10
France	8232.26	0.47	0.38	0.55
Germany	11761.27	0.60	0.53	0.66
Greece	9905.62	0.51	0.43	0.59
Hungary	8478.65	0.60	0.53	0.68
Iceland	8642.28	0.10	0.04	0.16
Ireland	8545.64	0.16	0.10	0.21
Italy	8373.22	0.53	0.45	0.61
Korea	5144.63	0.40	0.32	0.47
Luxembourg	9510.25	0.27	0.09	0.45
Mexico	7090.29	0.53	0.45	0.61
New Zealand	12057.62	0.17	0.12	0.23
Norway	10200.14	0.07	0.04	0.11
Poland	8975.21	0.59	0.50	0.68
Portugal	9068.37	0.36	0.28	0.44
Spain	7213.15	0.22	0.16	0.28
Sweden	8122.29	0.07	0.04	0.11
Switzerland	10423.20	0.42	0.36	0.49
United Kingdom	10017.19	0.31	0.26	0.36
United States	10826.89	0.27	0.20	0.34
Non-OECD countries				
Brazil	7586.31	0.46	0.39	0.52
Latvia	10264.05	0.29	0.21	0.38
Russian Federation	8170.58	0.33	0.26	0.39
PISA-plus countries				
Albania	10286.00	0.41	0.33	0.49
Argentina	10507.14	0.44	0.36	0.53
Bulgaria	11394.92	0.57	0.49	0.65
Chili	8366.54	0.51	0.42	0.59
Hong Kong	7349.67	0.50	0.42	0.59
Indonesia	5043.70	0.46	0.40	0.52
Israel	12094.30	0.47	0.39	0.56
Peru	11706.09	0.63	0.55	0.70
Thailand	6667.43	0.35	0.28	0.41
Macedonia	9044.18	0.46	0.34	0.58
Netherlands**	6973.92	0.47	0.37	0.57

* For calculation see annex 3
** response rate is too low to ensure comparability

As stated above, at the beginning of this section high between school variance indicates the degree of selectivity or segregation in a school system. The total between student variation on an achievement test in a particular country can be read as a measure of inequality of education. A large total between student variation indicates that an education system produces a lot of dispersion in actual learning outcomes; usually implying that a large proportion of students achieves at the low end of the score distribution. Further analyses of this distribution, for example by indicating which part of the student population is in the lowest percentile or quartile of the distribution can clarify this issue of inequality further. Most segregated and "unequal" are systems, which combine high total between student variation and high between school variation. Systems, such as Korea and the Netherlands, characterized by relatively low total between student variation and high between school variation, are selective in grouping students in schools, but manage to keep the total variation in achievement between limits. From an equity perspective school systems with high between school variation are still undesirable, since the selectivity is likely to be based on the socio-economic status of the students. School systems that have low between school variation and high total variation are probably systems with high degrees of internal tracking or streaming, leading to high between classes variation (compare the position of New Zealand in Tables A.2 and A.3). Most "equal" are school systems that combine low total between student variation and low between school variation. Sweden is a case in point.

These examples show that studying patterns of variance based on international comparative assessments provide additional information to comparing average achievement levels. More particularly these patterns provide information on the internal segregation and differentiation of school systems and corresponding implications for the distribution of learning outcomes.

Separating the effects of "given" background conditions and malleable school variables

In school effectiveness research it is standard practice to adjust student achievement scores for background conditions, preferably prior educational achievement in the same subject matter area or scholastic aptitude and, as a second choice, by adjusting on the basis of socio-economic status or minority background. Only after these adjustments have been made are the impact of malleable school variables tested. It appears, however, that even after these adjustments have been made, the "aggregates" or "composites" of the student background characteristics, like the school's average socio economic status, still explain a sizeable part of the between school variation.

The "net" effects[1] of schooling can therefore be attributed to two categories of variables: student composition and malleable school variables like leadership styles, school climate and instructional strategies. Compositional affects are likely to be thought of as "given" factors, while the factors that are malleable in the sense that they are seen as "handles" to improve the primary processes of schooling, teaching and learning. On further reflection, however, it is

[1] "Net" effect in the sense of student achievement adjusted for student background characteristics at the individual level.

clear that composition is also malleable, namely on the basis of overt admission or selection policies of the school. Or, influenced by selection processes by parents and students in the case of free school choice.
Databases of international comparative assessment studies provide an interesting source for assessing the relative impact of malleable school variables and school composition.

In Table A.4, cited from Scheerens and Visscher, 2004, the relative impact of these two categories of factors is shown for the three subject matter areas that were covered in the PISA 2000 study, reading literacy, mathematics and science.

Table A.4 (source Scheerens and Visscher, 2004) Partitioning of the between-school variance in reading, mathematical and scientific literacy by student background characteristics, school context variables, all school variables, and each of three groups of malleable school characteristics.

			Malleable school characteristics			
	Student background (%)	School context Variables (%)	All school variables (%)	School resources variables (%)	School climate variables (%)	School process variables (%)
Reading						
Overall	10.7	47.8	7.8	1.5	6.1	1.2
OECD	12.7	48.1	8.1	1.5	6.1	1.7
Mathematics						
Overall	23.2	30.9	7.8	1.5	6.0	1.3
OECD	26.3	29.6	8.9	1.4	6.7	1.9
Science						
Overall	25.9	29.8	7.4	0.9	6.1	1.2
OECD	28.6	29.5	8.2	1.1	6.4	1.7

In the figure, also cited from Scheerens and Visscher, 2004, these patterns are visualized for all the participating countries.

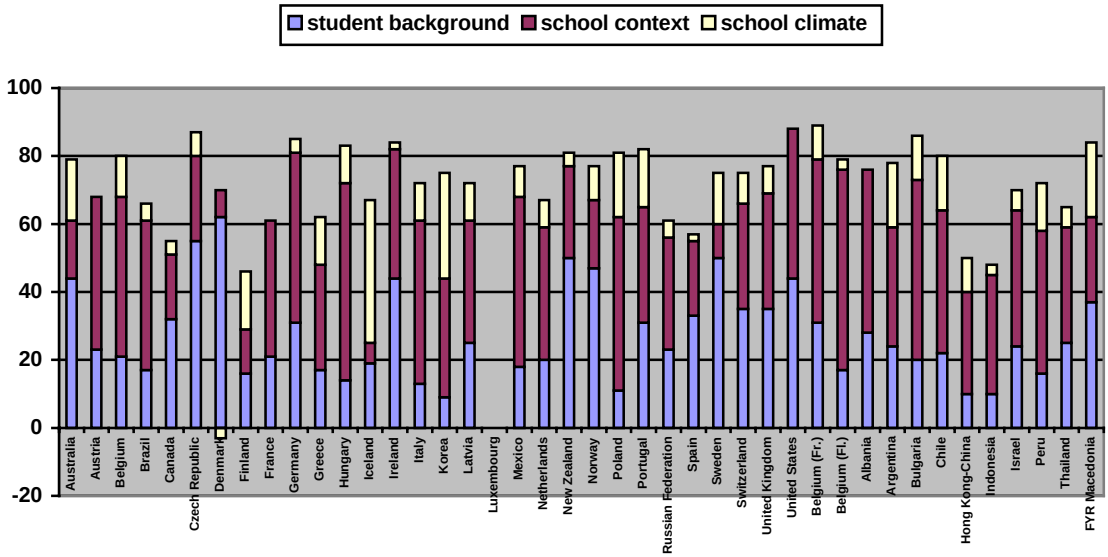


Figure A.1 Percentages of between school variance in reading literacy explained by student background variables, school context variables and malleable school variables

It is interesting to note that the between school variation is explained for a very large part by the student background variables and the school context variables (mainly the average socio-economic status of a school) and that only a relatively small part is explained by malleable school variables.

When we compare these results with similar analyses carried out in school effectiveness research studies, the balance between the impact of background conditions and composition on the one hand and malleable school conditions on the other is less extreme as in the case of the PISA data. In school effectiveness studies one is likely to find about 10% of the *total between student variation* explained by measured school variables, which, given a total between school variance component of 30, would be equal to explaining 30% of the between school variance.

The low estimates of the effects of malleable school variables that one usually finds in international comparative assessment studies, in comparison to school effectiveness research studies can be attributed to two causes:

- the usually rather superficial way of measuring school factors, namely by just using one or two questionnaire items to cover complex concepts;
- the lack of pre-test data in practically all international comparative assessment studies.

The impact of individual student level background conditions and their averages (compositional effects) on performance can be interpreted as indicators of (in)equity. Conditions of schooling are seen as less equitable to the extent that performance depends on the socio-economic background or the ethnicity of the students. The PISA data show that countries with a categorical system of secondary schools show larger impact of ses-related variables than do countries with a comprehensive system (Scheerens & Visscher, 2004).

A difficulty in making unequivocal comparisons between compositional effects on the one hand, and malleable school variables on the other is that these two categories of variables "overlap" in their impact on performance. Scheerens and Visscher (2004) estimated this "overlap" as a joint effect of both categories of variables. For many countries the joint effect is larger than the two unique effects. Interestingly enough countries differ considerably in the magnitude of this joint effect. Since the joint effect comes down to favorable conditions of schooling "going together" with a student population of the school that has favorable background characteristics the joint effect can be interpreted as another indicator of (in) equity. The cross-sectional nature of the data in a study like PISA 2000, precludes a sharper identification of the kind of selection processes that give rise to the joint effect of malleable an school composition variables. It might be the case that schools with better teaching conditions attract "better" students or that favorable characteristics of the students attract better teaching conditions.

Assessing the effectiveness of specific school, context and classroom characteristics

The effectiveness question considers the impact of specific school, school context and classroom characteristics on performance. International comparative assessment studies can address the effectiveness question to the extent that school and classroom variables are actually measured, usually on the basis of questionnaires administered to school directors, teachers and/or students. The fact that effects of these variables are being assessed in a multitude of countries provides the interesting possibility to establish whether "what works" in one country also works in the next. Stated in less popular terms this question refers to the generalizability of effectiveness enhancing conditions across countries.
A few illustrations will be provided, based on SIMS (the Second International Mathematics Study of the IEA, the IEA Reading Literacy Study, and PISA 2000).

In table A.4 significant associations of a set of school variables and mathematics achievement from SIMS are shown for 17countries (Source: Scheerens, Vermeulen & Pelgrum, 1989).

Table A.4: Predictor variables with significant positive (+) or negative (-) associations (5% level) with mathematics achievement, when the variance component model is analyzed by means of the VARCL-Programme

Country	Belgium (Fl.) 15	Belgium (Fr.) 16	Canada (B.C.) 22	Canada (Ont.) 25	Finland 39	France 40	Hong Kong 43	Hungary 44	Israel 50	Japan 54	Luxembourg 59	Netherlands 62	New Zealand 63	Scotland 72	Sweden 76	Thailand 79	U.S.A. 81	Total
Predictor variable																		
Fathers'occupation *(yfocci)*	m	m		+	+	+		+	m				+	+	+	+	+	9
Fathers' education *(yfeduc)*									+									1
Level of expected further education *(umoreed)*	+	+	+	+	+	+	+	m	+	+	+	+	+	+	+	+	+	16
Homework *(yahwkt)*			-				+	+	+	+		-	+		+	+		9
Teacher experience *(texpmth)*						+	+				+	+	+					5
Time spent keeping order *(tordert)*		-			-	-							-				-	5
Time spent on teaching *(tlistt)*	+						+					+			+			4
Teacher expectations *(ttop)*			+	+	+	+	+	+	+		+	+	+		+		+	13
Use of published tests *(tpubtst)*			+		+			+										3
Use of own tests *(towntst)*							-									+		2
Opportunity to learn *(totl)*	+	m		+		+	m		+	+		+	+	m	+		+	9
Class size *(klgrt)*	-	+			+	+						+	+	+	+			8
Urbanization *(saera)*		+					m	+				-						3
Number of woman teachers *(ssommf)*	+										-				-			3
Number of male teachers *(sallmm)*	+							+	-									3
Number of meetings *(smeet)*																		0

Note: School and classroom predictor variables are corrected for father's occupuation or father's education – when a predictor variable was not measured in a country this is indicated by the letter m.

The authors conclude that only a small number of school/classroom characteristics show a consistently positive association with mathematics achievement. These factors are: positive expectations of pupils' achievement (the variables *ymoreed* with an average association of .19 with achievement and *ttop*, average association of .22), and opportunity to learn (average association of .15). The authors (ibid. p. 797) go on to critically analyze these few positive associations. "The educational significance of the positive results might be challenged on conceptual and statistical ground. One could argue that associations of variables such as "positive expectations" and "opportunity to learn" with achievement, are something like a tautology. In the worst case, opportunity to learn could reflect the purposeful training of test items. "High expectations" might just as well be seen as the *effects* of high achievement rather than one of its *causes.*" They also conclude that variables that have received empirical support in the international research literature on school and instructional effectiveness, like frequent evaluation of students' progress, teachers' experience and "time on task" were found to have weak and/or inconsistent effects across countries.

Posthlethwaite and Ross (1992) followed a different approach in their analysis of the data from the IEA Reading Literacy Study. In each country they identified variables that significantly discriminated between the 20% highest and the 20% lowest scoring schools in the country. In this way they could produce a list of those variables that discriminated high and low effective schools in a certain number of countries. The relevance of the variables could thus be judged in terms of the number of countries in which a particular variable discriminated. The results are summarized in table A.6

Table A.6: *Teacher and school indicators discriminating effective and ineffective schools (top 15)* (Source: Postlethwaite & Ross, 1992)

Rank	**Indicator**	**No. of countries**
1	degree of parental cooperation	16
2	reading in class	17
3	no serious problems	18
4	urban-rural	14
5	school size	12
6	community resources	14
7	reading materials in schools	13
8	comprehension instruction	11
9	percent female teachers	14
10	classroom library	10
11	total teaching experience	11
12	school resources	13
13	student-teacher ratio	12
14	sponsor reading initiatives	13
15	literature emphasis	9

Scheerens and Bosker, (1997) re-analyzed this data set, using multi-level analyses. Their results with respect to overall effects across countries, using the total data set, were summarized as follows.

"Both context indicators *public/private* and *rural/urban* show a positive association with adjusted school effects in reading, showing advantages for private and urban schools. From the input indicators *class size* has a small, and meaningless, positive effect, and parental involvement has a clear positive effect (.08).

From the school process variables two achievement press variables (*focus on higher order problem solving skills* and *focus on reading*) have significant but small (.02) positive effects. The *consensus & cooperation* indicator has a significant but small (-.02) negative effect. The climate indicator shows a somewhat higher association (.04).

The other school process variables have estimated effects that are, statistically speaking, not discernable from zero. Of all teacher/classroom process variables only one has (an unexpected) negative effect: -.02 namely is the effect of time for reading"

And they conclude:

"All in all the model for the international data does poorly, with only 9 percent of unique variation between schools accounted for by the educational effectiveness variables" (Ibid p. 260).

A final illustration is based on PISA, 2000, source Scheerens and Visscher, 2004. After student achievement in reading literacy had been adjusted for student background conditions the following school variables appeared to have a significant association with performance when the whole data-set was used:

Figure A.2 School variables significantly related to reading literacy performance, after adjustment for student background characteristics (Source: Scheerens & Visscher, 2004).

SCHOOL RESOURCES VARIABLES
• school size • index of the quality of schools' educational resources • proportion of teachers with a third level qualification
SCHOOL CLIMATE VARIABLES
• index of disciplinary climate • index of teacher support **(-)** • index of teacher-student relations • index of students' sense of belonging tot the school • index op principals' perception of teacher-related factors affecting school climate (-) • index of principals' perception of student-related factors affecting school climate

SCHOOL PROCESS VARIABLES
• students' performance is considered for school admission* • transfer of low achievers to another school*

*) significant for OECD –countries only

When associations with unadjusted performance scores are considered (see the initial OECD report on PISA) considerably more school variables, such as school autonomy appear to be significantly associated with performance; effects that disappear when the proper adjustments are being made. Wößmann, (2000), incidentally reports a significant effect of school autonomy on the basis of an analysis of the TIMSS data set, using a country level model.

Willms and Somers, (2001) report findings that are more in line with the knowledge base on school effectiveness. Their analyses are based on UNESCO's *Primer Estudio Internacional Comparaivo (PEIC)* on 13 Latin American countries.
These authors conclude that the most effective schools are those with:
"1) high levels of school resources, including a low pupil-teacher ratio, more instructional materials, a large library, and well-trained teachers;
2) classrooms which are not multigrade, and where students are not grouped by ability;
3) classrooms where teachers are tested frequently"
4) classrooms and schools with a high level of parental involvement; and
5) classrooms that have a positive classroom climate, especially with respect to classroom discipline" (ibid. p. 439)

In conclusion it can be said that generally the results of associating school and classroom variables in international comparative assessment studies have been somewhat disappointing as far as the "global" studies or IEA and the OECD are concerned. Consistently smaller associations are found as in the case of national empirical school effectiveness studies. Moreover, consistency in certain variables being associated with performance across countries is also relatively disappointing.
The same methodological explanations could be given as the ones that were presented in the previous section: lack of longitudinally measured performance and relatively weak operationalizations of the process variables. At the same time part of the results might also be due to genuine differences between countries, or cultures. The PISA-re-analysis appears to point out that the school effectiveness variables that are known from the literature "work best" in traditionally English speaking countries. In these countries most of the empirical school effectiveness studies have been carried out as well. Nordic countries generally do very well in these international assessments but probably due to a somewhat different set of conditions, like the esteem for the teaching profession and the value education has in the society. Climate variables, also part of the school effectiveness heritage, work well in the Nordic countries as in countries with an Anglo-Saxon tradition.

Conclusion: making up the balance on the usefulness of international comparative assessment studies for answering questions about educational productivity and effectiveness

International comparative assessment studies are particularly useful for assessing the productivity of education systems, in terms of average achievement in a specific subject matter area or literacy domain. The validity of these kinds of outcome measures as compared to, for instance, attainment measures, have been discussed in the chapter on measuring quality. Countries can pick and choose the benchmarks they would like to use, to compare themselves: the international average, the score of a neighbor country or the average of the highest scoring country. As the illustrative data from PISA and TIMSS have shown large differences exist between the highest and the lowest scoring countries. For resources poor countries this might be problematic, because students might feel discouraged in not being able to do a substantial part of the items. At the same time it could be seen as important that such international comparisons can be made. A possible solution might be to expand the difficulty range in the sense of including sets of easier items for countries that are expected to score relatively low. If tests confirm to the assumptions of item response models, these easier item sets could then be vertically equated to the general international tests. International assessments for specific regions, like the PASEQ and SACMEC studies in Africa, and the *Primer Estudio Internacional Comparativo* in thirteen Latin American countries, have the advantage of being able to choose a more adapted difficulty level of the achievement tests, and include perhaps more ecologically valid items on the context of schooling in resources-poor countries.

Not only achievement *levels* such as the country averages are useful but also the patterns of *variability* that the score distributions of international comparative assessments show. As has been illustrated interesting conclusions can be drawn on the basis of the total between school variance, the proportion of the variance that is between schools, (usually indicated as the between school variance), and sometimes also the variance between parallel classes in one school. If the data can be broken down according to regions within countries, such analyses of the patterns of variability gain in relevance. Variability measures provide indications about the inequality among students in their achievement results, about the degree of segregation of the system of schools, and into practices like ability grouping and streaming within schools. In theory it would also be feasible to set benchmarks for keeping the different types of variability of an within schools systems within limits. Such benchmarks would speak to the equity interpretation of educational quality.

Comparing the impact of malleable school variables on the one hand and student background conditions and composition factors on the other, indicate the margins of control and change in education. On further reflection these different effects can be related to two different strategies to influence outcomes: productivity improvement on the one hand, and selection and admission policies on the other. The size of the composite effects, as was illustrated on

the basis of the PISA data set, may give rise to pay more attention to "selection management" and establishing fixed "quota" of students with specific background characteristics. It cannot be excluded that the impact of student background and compositional factors is overrated in international comparative assessment studies, because of weaknesses in the operationalization of the school variables. Besides, as was illustrated as well, the two types of factors overlap in their impact on achievement, which further complicates interpretation. In any case do international assessments provide the occasion to globally examine the margins of "malleability" in schooling, as well as the degree of dependency of results on student background characteristics and their aggregates. The latter providing an additional interpretation relevant for the equity perspective, implying that systems in which achievement results depend to a larger degree on "given" student background conditions like their socio-economic status, are considered to be less equitable than systems for which this association is lower.

The global international assessment studies from IEA and OECD have yielded relatively disappointing results with respect to confirming the effectiveness enhancing factors that are part of the school effectiveness knowledge base. This applies both to the size of the association of these variables with performance, after controlling for student background conditions, as to the weak consistency of the significance of the effects of these variables across countries. Regional studies, like the Latin American PEIC, however, do show results that are more in line with results of school effectiveness research studies. One way of improving the relevance of international comparative assessment studies to answer questions about educational effectiveness would be to invest more in measuring school factors and processes, using more extensive scales and perhaps also direct classroom observations. Another alternative would be to consider stand-alone school surveys and classroom observation studies to yield information on effectiveness enhancing process indicators. An example is the school and teacher survey in the countries united in the World Education Indicator Project of UNESCO, OECD and the World Bank.

References

Bishop, J. (1997) *The effect of National Standards and Curriculum-Based Exams on Achievement*. Cornell University. Center for Advanced Human Resource Studies.

Postlethwaite, T.N., & Ross, K.N. (1992). *Effective Schools in Reading. Implications for Educational Planners*. The Hague: IEA.

Scheerens, J., & Bosker, R.J. (1997). *The Foundations of Educational Effectiveness*. Oxford: Elsevier Science Ltd.

Scheerens, J., Vermeulen, C.J.A.J., & Pelgrum, W.J. (1989). Generalizability of school and instructional effectiveness indicators across nations. In: B.P.M. Creemers & J. Scheerens (eds.), *Development in school effectiveness research*. Special issue of the *International Journal of Educational Research*, 13(7), 789-800. Oxford: Pergamon Press.

Scheerens, J., & Visscher, A.J. (2004). *School factors related to quality and equity*. PISA thematic report. Paris: OECD (in press).

Willms, J.D., & Somers, M.-A. (2000). *Schooling outcomes in Latin America*. Report prepared for UNESCO-OREALC and the Laboratorio Latinoamericano de la Calidad de la Educación [The Latin American Laboratory for the Quality of Education].

Wöβmann, L. (2000). *Schooling resources, educational institutions, and student performance: The international evidence*. (Kiel Working paper No. 983). Kiel, Germany: Kiel Institute of World Economics.

Annex to:
REVIEW OF SCHOOL AND INSTRUCTIONAL EFFECTIVENESS RESEARCH
CONTRIBUTION TO CHAPTER 3 OF THE 2004 EFA GLOBAL MONITORING REPORT

ANNEX 2 FROM SCHOOL EFFECTIVENESS TO SCHOOL IMPROVEMENT

Jaap Scheerens, March, 2004

Introduction

The knowledge base on educational effectiveness provides tentative suggestions on what "generally" works in education. When the directly practical or political question is asked *how to make things work* accordingly, it will be clear that more will be involved than just disseminating the information on "what works" to policy-makers and practitioners (although this, in itself, is not a bad idea). It could be argued that the knowledge-base on school effectiveness provides direction to the *substance* of what could be tackled in education for the purpose of enhancing performance. Asking the question of how to initiate and steer the *process* of actual effectiveness improvement, however, may be seen as requiring the additional perspective of planned change in the education domain.

The action potential of the school effectiveness knowledge base

Before addressing various approaches within the overall field of educational change it is important to point at the action potential that integrated school effectiveness models intrinsically have. This action potential can be summarised in the following points:

- the school effectiveness knowledge base indicates malleable characteristics at various levels; which means that these characteristics are malleable by various kind of actors (central and regional officers, municipalities, school boards, school managers and school teachers);
- multi-level school effectiveness models, such as the one depicted in Figure.., assume influence across levels; for example the general notion that conditions at higher levels provide facilitating conditions as well as direct impacts on lower levels; this suggests that educational reform and improvement could (and maybe should) be tackled from various levels at the same time; e.g. central administration and school management;
- specific effectiveness enhancing factors are associated with specific modes or vehicles of improvement in education; the factors that work are associated with the curriculum (opportunity

to learn), management and leadership roles (educational leadership), school organisation (coordination and alignment), professionalisation of staff, monitoring, evaluation and feedback, teaching methodology, school community relationships and educational finance (incentives to achieve). This association between factors that works and modes or vehicles of educational change, suggest that stimulating processes that enhance effectiveness should probably be an integrated mix of activities using various modes, like curriculum reform, accountability requirements, professional development of school leaders and teachers and school development;

- the factors that "work" can be explained as specific instances of more general principals of management and control; in my earlier paper for the EFA Global Monitoring Team, I discussed synoptic planning, creating market mechanisms, retroactive planning, creating fit with contextual conditions, and self-reference in organisations as more general mechanism for change. A "meta-question" of effective schooling concerns the relative effectiveness of these mechanisms or levers for change. In the previous paper it was argued that the "cybernetic principle" involving evaluation, feedback, and possibly reinforcement holds a lot of potential, that coincides well with certain patterns of functional decentralisation and the perspective of the school as a learning organisation (Scheerens, 2003).

Part of this section of the earlier paper is repeated below, to briefly describe the various mechanisms or "levers for change".

Overview of theories and core mechanisms

Theory	*Core mechanism*
Rational control	Proactive structuring
Public choice	Market-mechanism
Retroactive planning	Cybernetic principle
Contingency theory	Fit
Theory of autopoietic systems	Self-organization

The rationality model that seeks to establish an instrumental and scientific approach to planning and control is well-known. Its imperative being, think before you act. Planning in terms of explicit goals and methods that are most effective and efficient in reaching these goals is still a predominant orientation, also in the field of educational reforms (compare the so-called logical framework approach). At the same time this model has been criticized enormously by scholars in the field of public administration, particularly for its stringent information demands and implicit assumptions of social harmony and consensus about social goals and means.

Public choice models take into account the fact that actors may not only pursue organizational goals, but also personal goals. The "royal road" to diminish the in-efficiency enhancing

tendencies is to install market conditions and competition in public sector organizations, including education.

Retro-active planning starts from the premise that action often precedes thinking, and that rational behavior might take the form of “rationalization” or rational reconstruction rather than pro-active structuring and planning. In more practical terms this approach suggests a change in the starting point of the planning cycle: start with an assessment of the current situation, identify strong and weak points, and plan and implement corrective, improvement oriented action.

Contingency-theory emphasizes the position that the effectiveness of organizational arrangements depends on situational conditions. Fitting approaches in teaching and instruction to characteristics of the target population, as in adaptive teaching, confirm to the contingency paradigm. The same could be said with respect to allowing schools a sufficient degree of autonomy to adapt to the local environment.

The theory of autopoietic systems (Maturana & Varela, 1980; Luhmann, 1995) emphasizes the importance of self-organization and self-reference in organizations. This theoretical approach might have important implications for a completely different view on issues of school effectiveness and school improvement. By emphasizing endogenous factors, and internal dynamics of organizations, it puts the whole issue of adaptation to the environment and external control on a different footing. In more practical terms the implications from this theoretical approach seem to be relevant to rethink issues of self-evaluation and feedback, the implementation of externally stimulated innovations (do innovations penetrate to the level of an organizations pattern of self-references?) and the whole issue of school autonomy.

The theoretical idea of retroactive planning in “learning organizations” is worth looking with a bit more detail. It coincides well with a pattern of centralization/ decentralization where processes are liberated and input and process control are being replaced by output control. This pattern is sometimes even indicated as the new management paradigm.
In more practical terms one could observe that it has the advantage of rooting reflection on the functioning of schools with an eye to school improvement, in empirical facts. As such it differs from pro-active approaches like school development planning that have often resulted in idealistic sounding documents that lead a quiet existence in various drawers of cupboards, and have little impact on what actual happens in schools.

The action potential of the school effectiveness knowledge base, laid out in these points, implies that it is possible to formulate reform agendas and improvement strategies that follow the logic of multi-level governance. More specifically, these may be worked out as strategies that are a combination of several factors. Substantive factors directly tackled (e.g. educational leadership);

vehicles or modes that are used to facilitate change in the factors in question (e.g. professional development); and mechanisms that are applied to bring about the actual change (e.g. synoptic rational planning). For an example, see Scheerens, 2000, chapter 4.

Systemic reform

Terms like "systemic reform" and "restructuring" are used to refer to changes in the institutional and organizational infrastructure of a country's education system.
New institutional economics, cf. North (19..) emphasizes the importance of institutions; where institution are "the rules of the game" or basic legislative arrangements. "Marriage" and property rights, are frequently mentioned examples of institutions. In developing countries quality assurance or quality improvement might start with an analysis of the institutionalization of basic arrangements in the education system, like: fixed teacher salaries; responsibilities of head teachers, official working time, time tables etc.
Organizational infrastructure might have to be scrutinized as well. The "organizational capacity" of the country's Ministry of Education, for example, might be analyzed and found to be in need of improvement (Orbach, 1998).
Questions about organizational capacity of an educational system first of all regard the issue of whether core functions have an "organizational home" in the system. For example, initiating a national assessment is the more of a heavy task when there exists no organization that has specialized in the development of educational achievement tests in the country. The same applies when external supervision of schools is considered at a fairly large scale and the country has no educational inspectorate.
Further criteria in determining the organizational capacity concern the well-functioning of organizations in terms of effective leadership, ability to mobilize financial, material and human resources and appropriate work practices (ibid).

The division of decision-making authority across the administrative levels of the educational system has both institutional and organizational significance.
In Western countries "restructuring" and systemic reform are usually focused at decentralization of decision-making authority and creating arrangements for accountability. Sometimes these two major dimensions are combined in certain patterns or arrangements that gear decentralization to accountability arrangements. So called "Performance-based approaches to large-scale reform" (Leithwood et al., 1999) form a case in point.

Decentralization and functional decentralization

The issue of decentralization acquires considerable more nuance and practicality when it is being recognized that it is possible to decentralize in particular domains of decision-making, while

doing rather the opposite in other domains. This notion comes close to what others have called “functional decentralization” (Bray, 1994). The concept of functional decentralization will be explained in more detail by referring to an instrument and data collection procedure that has been developed in the context of OECD.

The OECD-INES procedure to measure “locus of decision making” distinguishes *three* facets of the rather crude distinction between centralisation and decentralisation:

- the tier or administrative level where a decision is taken; this dimension was referred to as the *locus of decision-making*;
- the amount of discretion, or the degree of autonomy of decision-making at a particular administrative level; this facet was called the *mode of decision-making*;
- the particular element of educational administration a decision belonged to; this facet was referred to as the *domain of decision-making*.

These three facets can be related to existing categorisations in the relevant literature, although the use of central concepts is by no means consistent among authors and publications. Our three-dimensional conceptualisation is compared to the terminology as clarified by Bray (1994, p. 819) in an analysis of alternative meanings of centralisation and decentralisation.

The distinction between levels confirms to the concept of *territorial decentralisation*, defined as "the distribution of powers between different tiers of government". In the operationalization of this dimension we distinguished four tiers, to be further described in the section on methods.

Degrees of autonomy in decision making at a particular level are reflected in terms that refer to an increase in discretion. Again following Bray, *deconcentration*, *delegation* and *devolution* are modes of decision making in which an increased amount of decision-making authority resides at a lower level.

"*Deconcentration* is the process through which a central authority establishes field units, staffing them with its own officers".

"*Delegation* implies a stronger degree of decision making at the lower level. However, powers in a delegated system still basically rest with the central authority, which has chosen to "lend" them to a local one".

"*Devolution* is the most extreme form of decentralization. Powers are formally held by local bodies, which do not need to seek approval for their actions" (ibid, p. 819).

In the operationalization of this continuum of increasing autonomy, these abstract definitions were avoided and respondents were asked to indicate whether decisions could be taken within the framework determined by a higher level, in consultation with a higher level or in full autonomy.

In order to determine elements or *domains* of educational administration, many categorization schemes are available in the literature (e.g. James, 1994; Winkler, 1989; Bacharach et al., 1990; Rideout and Ural, 1993). The common core of these categorizations are three main areas:

a) an educational domain (goals, methods, curricula, evaluation procedures);
b) an organizational, managerial and administrative domain (including human resource management, groupings and assignment and foundational regulations);

c) a dimension concerning finance and the way financial resources are applied.

In the operational classification that we chose four main categories were used, by splitting up area b (organisational) into two domains "planning structures" and "human resources", and including areas a and c.

The distinction between domains of decision-making in educational systems bears some resemblance to Bray's use of the term "functional decentralisation" as cited from Rondinelli. "Functional decentralisation refers to the dispersal of control over particular activities" (Bray, 1994, p. 819). From the examples that he provides, however, it is not clear whether in functional decentralisation an exhaustive set of domains of educational decision-making is referred to, as is the purpose of the categorisation schemes cited above. The common denominator is the recognition that educational systems may be centralised in some domains of decision-making but not in others. The conclusion is therefore that in a somewhat liberal use of the term our distinction between domains of educational decision-making can be considered as a form of functional decentralisation.

To learn more about educational decision-making in OECD countries and to systematically compare decision-making processes across countries, an instrument was developed that examined the locus of decision-making in four important domains. As stated above, these domains were: (1) the organization of instruction; (2) personnel management; (3) planning and structures; and (4) resource allocation and use. Within each of these four domains, between seven and 15 decisions were examined. In the domain entitled, “organization of instruction,” for example, the instrument focused on decisions about such matters as textbook selection, grouping of pupils for instruction, and assessment of pupils’ regular work. In “personnel management,” questions were asked about hiring and dismissal of teachers and other school staff, duties and conditions of service, and the setting of salary schedules. In “planning and structures,” the focus was on creation and abolition of schools and grade levels, the design and selection of programs of study, course content, and policies regarding credentials. Finally, in the area of “resource allocation and use,” the instrument focused on decisions about the allocation of resources for staff and materials, and the use of financial resources for these purposes.

Each of the questions in the instrument was designed to identify the level at which decisions are made in the governmental system (the “level” of decision making) and the way decisions are made (the “mode” of decision making). Six “levels” of decision-making were set out in the instrument. These include the following: (1) central government; (2) state governments; (3) provincial/regional authorities or governments; (4) sub-regional or inter-municipal authorities or governments; (5) local authorities or governments; and (6) schools. Three “modes” of decision-making were examined in the instrument. Decision could be made by an authority (1) autonomously, (2) within a framework established by another level within the system, or (3) in consultation with other levels in the system. Based on the instrument, it was possible to

determine how centralized or decentralized decision was overall, in each of the four domains, and for individual education decisions.

Accountability; evaluative capacity and incentive based policies

In general terms, accountability refers to holding public institutions and services responsible for the quality and output of their performance. Glass (1972) states that accountability involves several loosely connected strands: "disclosure concerning the product or service being provided; product or performance testing; and redress for poor performance (Glass, 1972). The third element implies that accountability is not just a matter of providing and judging information but at least also "foreshadows" actions by competent authorities in the sense of sanctions or rewards.
The first element – disclosure- requires that educational units, schools, in particular, provide information on their service provision, and make themselves "open" for external inspection and review. The second element distinguished by Glass stipulates that output and product information should be part of the disclosure on service provision and functioning. The third element emphasises that testing and review have implications in the sense of rewards and punishments for organisations. This relates accountability to incentive-based policies, like merit pay of teachers and output related financing of schools.
Types of accountability are distinguished on the basis of who, or rather which kind of unit or stakeholder, is supposed to use the information that is disclosed by schools and teachers, and also who is supposed to apply the sanctions.
Elmore and Associates (1990) differentiate three "theories" of accountability on the basis of this question: who uses the information. They distinguish three types:

- technical accountability, in which administrative units are supposed to take decisions on the basis of scientifically sound achievement measurements;
- the client perspective, in which the clients of education, like the parents of the pupils, "vote with their feet" in context of free choice of schools;
- the professional perspective; in which feedback on performance is basically used for professional development. "Accountability is, therefore, to be accomplished by deconstructing and reconstructing the meaning of schooling, collaborative planning, and co-operative teaching and learning" Elmore and Associates, 1990, cited by MacPherson, 1990, p. 7).

In my opinion only the two first forms can be seen as types of accountability. The "professional perspective" lacks the third element in Glass' basic definition, namely the application of rewards and sanctions. Moreover, what Elmore and Associates refer to as the professional perspective on accountability comes closer to the notion of "organisational learning" and the teacher as a reflective practitioner, as distinguished in the classical work of Argyris and Schön (1974). When specifying the professional perspective further, MacPherson also uses the term "empowerment"

of teachers, which is more closely associated with school-based and school initiated approaches to school improvement (see the text on school improvement approaches further on).

In order to make the two “real” forms of accountability work, systems should have *evaluative capacity*, that is structural and technical facilities to realise the kinds of empirical disclosure and performance testing that accountability requires. Scheerens (2000, informal paper for the OECD) mentions the following issues for assessing the evaluative capacity of education systems:

a) Availability of a legal framework that enforces types of educational evaluation?
   * If yes, specify the legal or semi-official (e.g. policy-plans or brochures from the Ministry of Education) requirements for:
     - external evaluation
     - internal evaluation

b) Does the system have an inspectorate? Tendencies in inspection.
   * If yes, describe
     - how the inspectorate is anchored in the decision-making structure
     - tendencies in policies towards and within the system of inspection
   * If not, how is the inspection function executed?

c) Is there a national (or above the school) curriculum, national standards, national assessment programme? In the case of Spain: how is the implementation in the regions?

d) From the idea that evaluation is in a sense technology-driven, give a short indication of the state of development of Educational Evaluation.
   * Number of Faculties of Educational Science;
   * Number of Research Institutes;
   * Number of Researches
   * Budget

e) Existence of specialized departments or institutes for educational testing and evaluation?
   * Give a brief description of the infra-structure of institutes on evaluation.

f) Evaluation culture?
   * Is there a special policy with respect to internal or external evaluation?
   * Characterize the general attitude of schools towards external and internal evaluation in terms of, for instance, cooperative, resistant, initiating, ..

g) Which other regulatory mechanisms, as compared to evaluation, are of particular relevance in the country?
   e.g. - mechanisms to select teachers and pupils
   - professionalization of teachers
   - financial input control

h) What is the state of affairs concerning debates on possibly undesired consequences of evaluation procedures for external or internal evaluation? E.g. political bias, resistance, 'red tape', undesired side-effects of hard competition between schools

With respect to the third defining element, the application of rewards and sanctions, which can be brought under the heading of incentive based policies, research shows that there are often considerable limitations. When it comes to technical or administrative accountability reviewers usually have to conclude that few examples of straightforward decision-making seem to exist. Chibulka and Derlin, (1995) in their review of systems of school performance reporting, for example, say that "school performance reporting (SPR) is not considered very important by policy-makers or the general public". They conclude that it has not been demonstrated at all that "SPR can become a potent, effective policy-lever". Similar reservations have been based on empirical studies of the use that parents make of school performance information in choosing a school for their children (Bosker & Scheerens, 1999). Nevertheless there is evidence that accountability raises actual student achievement. This is expressed, for example, in the following study results:

Bishop, 1997: "Countries and Canadian provinces with standards and assessment based reforms outperform other countries at a comparable level of development"

Rand News, July 25, 2000: "The most plausible explanation for the remarkable rate of math. Gains by North Carolina and Texas is the integrated set of policies involving standards, assessments and accountability that both states implemented in the late 1980s and early 1990s

The most plausible answer to the question in what way accountability works, seems to be that the evaluation and feedback of performance related information stimulates schools to be more result oriented and targeted towards the attainment of the outcomes that are actually measured. Critics to the accountability movement say that gains in test result are caused by "teaching to the test" and related strategic behavior (cf. Sacks, 1999).

Mixed patterns of decentralization and accountability provisions: the example of performance-based approaches to large-scale reform

Letihwood, Jantzi, and Mascall (1999) state the following properties of the "performance-based approach":

1. A centrally determined, unifying vision, and explicit goals for student performance, based on the vision.
2. Curriculum frameworks and related materials for use in accomplishing the goals set for students.
3. Standards for judging the quality of degree of success of all students.
4. Coherent, well integrated policies that reinforce these ambitious standards.
5. Information about the organization's (especially the students') performance.

6. A system of finance and governance that devolves to the local school site responsibility for producing improvements in system and student performance.
7. An agent that receives the information on organizational performance, judges the extent to which standards have been met, and distributes rewards and sanctions, with significant consequences to the organization for its success or failure in meeting specified standards.

Leithwood and his co-authors evaluated the impact of five performance-based reform projects (in Kentucky, California, New Zealand, Victoria (Australia), and Chicago) and concluded that only Chicago had demonstrated significant increases in student achievement. They also found that these achievement gains only occurred during the last three of the ten years the program was analyzed. According to Fullan (2000, cited by Hopkins, 2002) during the first six years of the program "the system operated in decentralized fashion with little functional contact between schools and the district. In other words too little structure characterized the operation". During the latter years of the program "five extra district-level functions were developed", and these might explain why students this better during the last years of the program that were considered in the analyses:

- policy making increasingly supported decentralization
- there was a focus on local capacity building
- a system of rigorous accountability was introduced
- innovation was stimulated
- external support networks were established

(Hopkins, 2001, p. 3)

Combined arrangements of functional decentralization and accountability that appear to be successful are characterized by centralization on the curriculum and assessment dimension and increased autonomy in areas like personnel management and resource management at school level. "A micro-economic student-level estimation based on data [TIMSS] from 39 countries reveals that positive effects on student performance stem from centralized examinations and control mechanisms, school autonomy in personnel and process decisions.." Wößmann, 2000.
The example of the Chicago reform program points the attention at two other dimensions that co-determine success:

- pronounced vertical coordination between higher administrative levels and the school level;
- taking into consideration and stimulating local capacity.

Local capacity building has always been one of the main issues in school improvement. School improvement being considered as a more school-based approach to educational change and innovation as compared to systemic reform as discussed in this section.

School improvement

School improvement as a field of academic study is seen as a specific branch of the study on educational change. As will be pointed out further on, in some applications it is explicitly related to the school effectiveness knowledge base, and, in still other applications, its insights are also combined and integrated in perspectives on systemic reform.

Matthew Miles' overview of the development of this field of study, in the period between the mid 1950's and the mid 1990's provides the flavour of what the field represents. Miles discusses ten consecutive school change strategies:

1*) Training for group skills*; i.e. teaching school people fundamental skills of group behaviour. Such skills were considered to be of key importance for developing co-operation, "process analysis" and self-reflection for school teams and also as a subject to be addressed in classroom teaching.

2) *Innovation, diffusion and adoption*; which came to the fore during the 60's. This was the period of programmed instruction and the idea of "teacher proof" curriculum materials; i.e. curriculum materials and teaching methods that were specified to such a degree that bad teaching could not spoil the deliverance to students. And even when it was recognised that teaching material required an "interpretation" by teachers, for example in adapting to local conditions, the criterion for successful implementation was coined in terms of the "fidelity" to the externally determined "script". Educational change experts like Miles, at first, thought of temporary systems, like task forces, that were more actively involved in interpretation, and later on developed concepts on more active adaptation and re-creation of externally induced innovations.

3) *Organisational self-renewal.* Following developments in industry that came under the heading of "organisational development", during the late sixties the school as an organisation was increasingly seen as the object and the agent of change. The aim was "to induce organisational self-renewal through tactics of training, process consultation, data feedback, problem-solving and structural change" (Miles, 1998, p. 48).

4) *Knowledge transfer.* In this area the more simplistic expectations of the Research Development and Dissemination (RDD) strategies were challenged, again (as in point 2) emphasising active reconstruction at school level of the knowledge that was offered from outside. Capacity building was seen as a necessary prerequisite of good knowledge transfer.

5) *Creation of new schools.* In this section Miles writes about the phenomenon that many new and alternative school projects came into being in the 1960's an 70's in the USA. He concludes that by analysing some of these, he learned that "good new schools can be created, but that the task is very demanding, more complex than expected, and requires assistance and political protection" (ibid, 50).

6) *Supported implementation.* In the late 70's the "passive" idea of adoption of externally induced change had been abandoned, and instead, implementation was being seen as a longer term process of "adaptation". Adaptation requiring that schools develop coherence and meaning to external change initiatives. Based on his experiences of assisting some large projects that recognised this

implementation perspective Miles concluded that "continued assistance" throughout the implementation process was of major importance.

7) *Leading and managing local reform.* According to Miles during the eighties there were a lot of local initiates "many of them pushing hard on effective schools and effective teaching programs". On the basis of studying successful projects, Miles and his colleagues identified characteristics of success local reform projects. He summarises as follows: "The ideas of vision-building, pressure and initiative taking, and assistance have already been outlined. The idea of *empowerment* is an extension of the concept of *legitimacy for planning and action,* indicating in sharper terms that we found reform success closely associated with the presence of a cross-role planning team with clear decision power over change-related matters (such as project budgets, staff development, staffing patterns, and related time)". He goes on to say that he found three variables that were tied to successful local reform. The notion that the planning style was "evolutionary" rather than "architectural". He describes evolutionary planning as "a journey in the service of an evolving, increasingly shared vision". Secondly he found that successful schools were good at resourcing and problem coping (the slogan: "problems are our friends").

8) *Training of change agents.* As concluded earlier, despite the importance of local initiative, school change is usually in need of external support and facilitation, according to Miles. Support, particularly with respect to the change process. In this context he identified two major characteristics of successful support. "*Developing trust and rapport.* A great deal seems to depend on a change agents' ability to develop a strong, supportive, contractually clear relationship with specific "clients" –groups and individuals involving in change efforts". The second characteristic of successful support is organisational diagnosis, a data driven "understanding" of schools as organisations.

The last two strategies that Miles mentions are in fact about integrating school change in system wide reform initiatives: "managing systemic reform" and "restructuring schools". Murphy (1993) states that "restructuring" in the USA usually has four main strategies for reorganising education: providing choice and voice for parents, school-based management, teacher empowerment, and teaching for understanding. The latter refers to a constructivist orientation to teaching and learning.

Taking in consideration other seminal contributions to the conceptualisation of school improvement, as those by Fullan and McLaughlin and Skillbeck, published in the "International Handbook of Educational Change" (1998) edited by Hargreaves, Lieberman, Fullan and Hopkins, the following can be seen as the key principles of this orientation to educational change.

a) The school is the focus of educational change. This means that schools should be analysed as organisations, seen in their local contexts and harbouring the major agents of change, namely teachers.
b) A strong emphasis on the process dimension of educational change.
c) The importance of school based "implementation" in the sense of active adaptation or "co-invention" of externally induced changes.

d) A human relations approach to educational change influenced by group dynamics and the idea of teacher "empowerment", capacity building and overcoming professional isolation of teachers. The "counseling" approach of external change facilitators perhaps also fits in this tradition.
e) An evolutionary "bottom up" view on educational planning and curriculum development.

Within the scientific community active in this field quite a range of emphases can be discerned. These vary from authors like Mitchell and Sackney (2000), who provide a post-modernist view on school improvement and are strongly opposed to accountability and other "mechanistic" approaches, to authors like Reynolds and Hopkins, who relate school improvement to the school effectiveness research in emphasising learning and learning outcomes. Still other contributions (e.g. Leithwood et al., 1999, and Hopkins, 2001) integrate school improvement approaches and conceptualisations of systemic reform.

A major break-through in this field is the work of Slavin, who has proposed a "third" way, in addition to the school improvement approach and systemic reform. (Slavin, 1996, 1998). The characteristics of the school improvement approach as described in the above are summarised by Slaving under the heading of "organisational development models". "Perhaps the dominant approach to school-by-school reform is models built around well-established principles of organisation development, in which school staffs are engaged in an extended process of formulating a vision, identifying resources (such as external assistance, professional development, and instructional materials) to help the school toward its vision, and often locating "critical friends" to help the school evaluate and continually refine its approaches". Of this approach Slavin says that it is time consuming and expensive. Moreover, he claims that it is only effective for schools that already have a strong capacity for change. "Such schools are ones in which staff is cohesive, excited about teaching, led by a visionary leader willing to involve the entire staff in decisions, and broadly aware of research trends and ideas being implemented elsewhere." (p. 1303). Such schools he describes as "seed" schools. A second category of schools Slavin describes as school who would like to do a better job, but do not perceive the need of the capability to develop new curricula. According to his categorisation these are school with good relations among staff and leadership, a positive orientation toward change, and some degree of stability in the school and its district. Finally, as a third category, he refers to schools " in which even the most heroic attempts at reform are doomed to failure. Trying to implement change in such schools is like trying to build a structure out of sand." (ibid 1303). Accordingly he refers to these schools as "sand" schools.
School improvement of the organisational development kind (as we have seen the predominant perspective on school improvement) is considered only feasible in "seed schools", which he estimates at 5% of all schools in the USA. Sand schools, also about 5% of all schools would require fundamental changes before they can support any type of school change. The overall majority of schools, according to Slavin, are the brick-schools and they could most efficiently benefit from what he calls *comprehensive reform models.* His own "Success for All" program is an example.

Comprehensive reform models provide schools with specific student materials, teachers' manuals, focused professional development, and relatively prescribed patterns of staffing, school governance, internal and external assessment, and other features of the school organisation. It should be marked that "Success for All" is one of the few improvement projects that has been thoroughly empirically evaluated and has shown to be successful (Slavin, 1996, Scheerens & Bosker, 1997). Similar successes have been reported by Stringfield and others (1995) presenting the idea of schools as "high reliability organisations".
It is interesting to note that Slavin's conception (and also its actual realisation in "Success for All") of Comprehensive Reform Models, seems to have returned full circle to the point where, according to Miles, the school improvement movement started its human relations/implementation approach in the 1950's. Namely the discussion on the applicability of externally developed pre-structured innovation programs and curriculum material. The fact that there is clear evidence that this approach works is revolutionary, and puts a question mark behind the efficiency of forty years of educational innovation based on the less directive, bottom up, social psychological, organisational development approach to school improvement. The question of efficiency seldom being raised from within this tradition so eloquently described in Miles ten strategies for school change.

Conclusion: Integrating systemic reform and school improvement utilising the school effectiveness knowledge base

Hopkins, 2001, provides in interesting case in which a systemic reform approach is augmented by insights from the school improvement perspective, namely by investing more efforts in local capacity building.
Such integrative conceptualisations seem to be very relevant for quality enhancement programs in developing countries.

But before discussing integrated reform strategies further it is important to address a more basic issue for quality enhancement in the education sector of developing countries. For these countries creating pre-conditions for reform and quality enhancement seems to be of crucial importance. Two notions of creating pre-conditions have been touched upon in the review of the literature on educational change. The first concerns institutional and organisational conditions at system level. The second is identified "by default" (because not further specified) in the citations of Slavin's work, particularly his conception of "sand" schools. Of these he says that such schools "require fundamental changes" before they can support any type of school change" (p. 1304). As in the case of basic institutional and organisational structural arrangements at system level, these pre-conditions for change at school level could be seen as pre-requisites that would need to be addressed before any other more ambitions improvement efforts could be started. The school effective research literature, particularly the part of this literature that is based on studies carried out in developing countries, offers some suggestions of what such basic pre-requisites of further quality enhancement

might be. Facilities of buildings and classrooms, resources for teaching and learning, a certain level of teacher training, clarity about curriculum objectives and a degree of stability in staffing and school functioning could be mentioned as research supported requirements.

The educational effectiveness literature has started to show some results on effective patterns of systemic reform. The results so far support standard based examination and evaluation capacity as well as a pattern of functional decentralisation that features centralisation in the curriculum domain and autonomy in other domains of school management. When comparing the two strategies for school improvement the "bottom up" organisational development approach (Muijs and Reynolds speak of "the ownership paradigm") and the externally guided approach (Slavin's Comprehensive Reform Models), it seems that in most situations in developing countries the latter would be more effective and efficient. Such an approach coincides well with systemic reform that centralises the curriculum domain.

References

Argyris, C., & Schön, D.A. (1974). *Theory in practice: increasing professional effectiveness*. San Francisco: Jossey-Bass.

Bacharach, S.B, Bamberger, P., Conley, S.C., & Bauer, C. (1990). The Dimensionality of Decision Participation in Educational Organizations: The Value of a Multi-Domain Evaluative Approach. *Educational Administration Quarterly, 26*(2), 126-167.

Bishop, J. (1997) The effect of National Standards and Curriculum-Based Exams on Achievement.. Cornell University. Center for Advanced Human Resource Studies.

Bosker, R.J., & Scheerens, J. (1999). Openbare prestatiegegevens van scholen; nuttigheid en validiteit. *Pedagogische Studiën, 76*(1), 61-73.

Bray, M. (1994). Centralization/decentralization and privatization/ publicization: conceptual issues and the need for more research. In W.K. Cummings & A. Riddell (eds.), *Alternative Policies for the Finance, Control, and Delivery of Basic Education* [pp. 817-824]. Special issue of the *International Journal of Educational Research, 21*(8).

Chibulka, J.G, & Derlin, R.L. (1995). State educational performance reporting policies in the U.S.: accountability's many faces. *International Journal of Educational Research, 23*(6), 479-492.

Elmore, R.F. and Associates,. (1990) *Restructuring School; The Next Generation of Educational Reform.* San Francisco: Jossey Bass

Fullan, McLaughin & Skillbeck (1998). In Hargreaves, Lieberman, Fullan & Hopkins (Eds.), International handbook of Educational Change.

Glass, G.V. (1972). The many faces of educational accountability. *Phi Delta Kappan*, 53, 636-639.

Hopkins, D. (2002). Educational Innovation: Generic Lessons Learned from (a) Regional Practice. Keynote Paper Presentation at the DECIDE International Seminar, University of Twente the Netherlands

James, E. (1994). The public-private division of responsibility in education. In W.K. Cummings & A. Riddell (eds.*), Alternative policies for the finance, control, and delivery of basic education*. Special issue of the *International Journal of Educational Research*, *21*(8), 777-783.

Leithwood, K., Jantzi, D., & Mascall (1999). Large-scale Reform. What works? Ontario Institute for Studies in Education, December, 1999.

Luhmann, N. (1995). *Social systems*. Stanford, Ca.: Stanford University Press.

MacPherson, R.J.S. (1990). *Educative Accountability: Theory, Practice, Policy and Research in Educational Administration.* Oxford: Pergamon.

Maturana, H.R., & F.J. Varela (1980). *Autopoiesis and Cognition*. The Realization of the Living. Dordrecht/Boston/London: D. Reidel Publishing Company.

Miles, M., (1998) Finding Keys to School Change: A 40-Yeat Odyssey. Hargreaves, A., Liebermann, A., Fullan, M., And Hopkins, D. *International Handbook of Educational Change.:* Dordrecht, Boston, London: Kluwer

Mitchell, C., & Sackney, L. (2000). *Profound improvement: building capacity for a learning community*. Lisse: Swets & Zeitlinger.

Murphy, J. (1993). Restructuring Schooling: the equity infrastructure. School Effectiveness and School Improvement, 4(2), 111-130.

North, D.C., (1990) *Institutions, Institutional Change and Economic Performance.* Cambridge, U.K. Cambridge University Press

Orbach, E. (1998) How to Assess the Capacity of Client Government to Implement an Education Development Project. Washington: World Bank.

Rand News Release (July 25, 2000) Explaining achievement gains in North Carolina and Texas.

Rideout, W.M., & Ural, P. (1993). Centralized and decentralized models of education: comparative studies. Development Bank of Southern Africa. Centre for Policy Analysis.

Sacks, P. (1999) *Standardized Minds.* Perseus Books. Cambridge, Massachusetts

Scheerens, J. (2000). *Improving school effectiveness.* Paris: UNESCO, IIEP, Fundamentals of Educational Planning series no. 68.

Scheerens, J. (2000a) How to Measure the Evaluation Capacity of National Education Systems. internal document Paris: OECD.

Scheerens, J. (2003). Conceptual framework for the World Education Indicator survey on schools and teachers. Paper for the OECD.

Scheerens, J., & Bosker, R.J. (1997). *The Foundations of Educational Effectiveness*. Oxford: Elsevier Science Ltd.

Slavin, R.E. (1996). *Success for all*. Lisse: Swets & Zeitlinger.

Slavin, R.E. (1998). Sands, Bricks, and Seeds: School Change Strategies and Readiness for Reform. Hargreaves, A., Liebermann, A., Fullan, M., And Hopkins, D. *International Handbook of Educational Change.:* Dordrecht, Boston, London: Kluwer

Stringfield, S.C., Bedinger, S. & R. Herman (1995). Implementing a private school program in an inner-city public school: processes, effects, and implications from a four year evaluation. Paper presented at the ICSEI Conference, Leeuwarden, the Netherlands.

Winkler, D.R. (1989). *Decentralisation in education: an economic perspective*. Washington:The World Bank, Population and Human Resource Development.

Wößmann, L. (2000) *Schooling Resources, Educational Institutions, and Student Performance: The International Evidence*. Working paper no. 983. Kiel, Germany: Kiel Institute of World Economics.